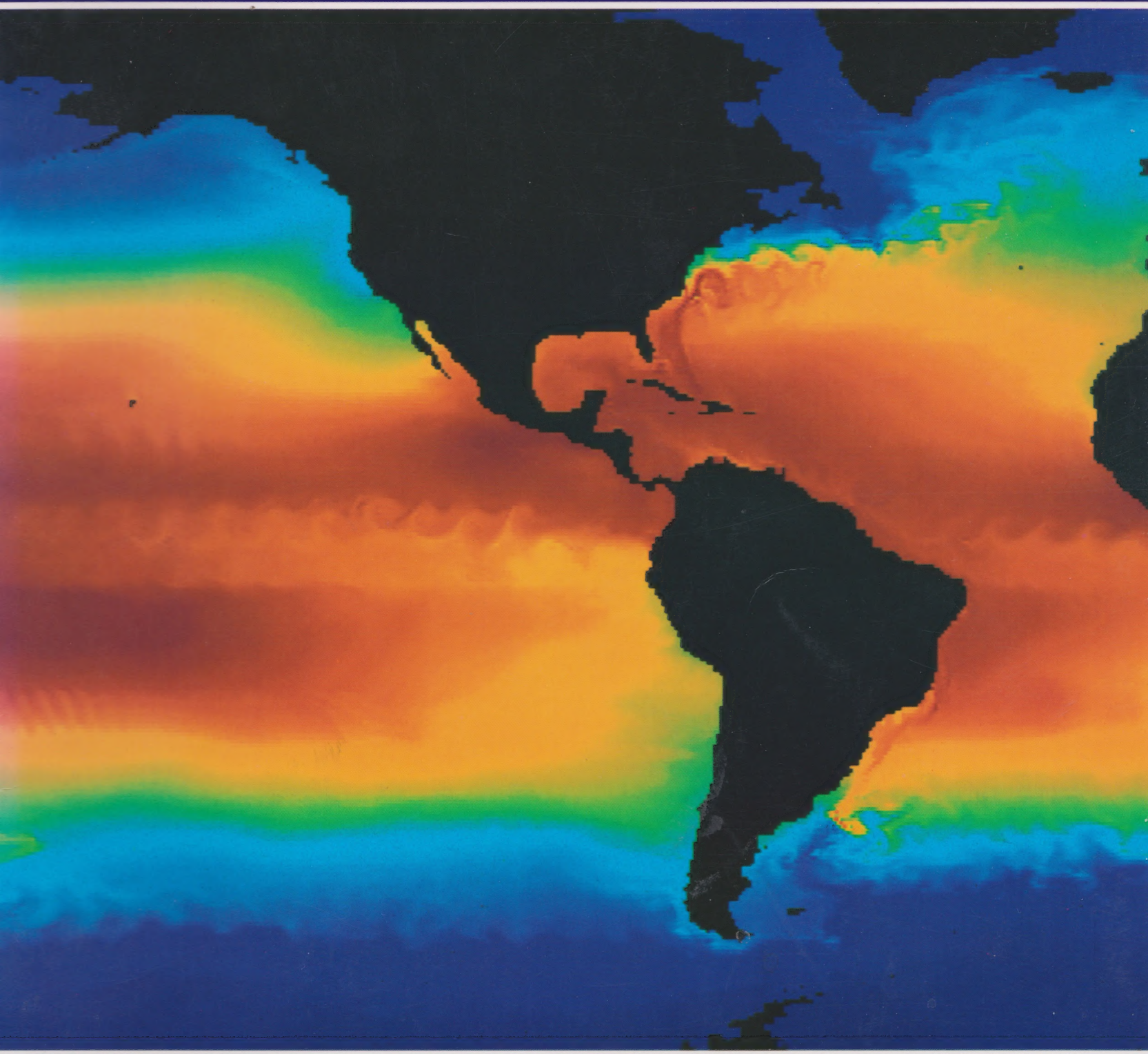
What's Happening in the Mathematical Sciences

Volume 1 • 1993

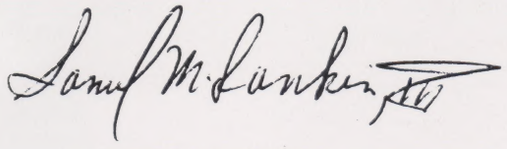


American Mathematical Society

Introduction

Welcome to the inaugural issue of *What's Happening in the Mathematical Sciences*! To be published annually, *What's Happening* surveys some of the important developments in the mathematical sciences over the past year or so. Mathematics is constantly growing and changing, reaching out to other areas of science and helping to solve some of the major problems facing society. Here you can read about the development of a mathematical model of the human heart, the solution to a longstanding mathematical problem about the way a drum's shape affects its sound, and the contributions mathematics is making to the solution of environmental problems.

What's Happening in the Mathematical Sciences aims to inform the general public about the beauty and power of mathematics. The American Mathematical Society is pleased to present this new publication. We hope you enjoy it!



Samuel M. Rankin, III
AMS Associate Executive Director
and Director of Publications

Cover Illustration. A group of scientists at Los Alamos National Laboratory have developed a mathematical model of ocean dynamics for massively parallel computers that they hope will improve understanding of the role of oceans in global climate change. The colors in this computer-generated picture indicate sea surface temperature from cold (blue) to warm (red). Figure courtesy of Richard Smith, John Dukowicz, and Robert Malone at Los Alamos National Laboratory.

Contents

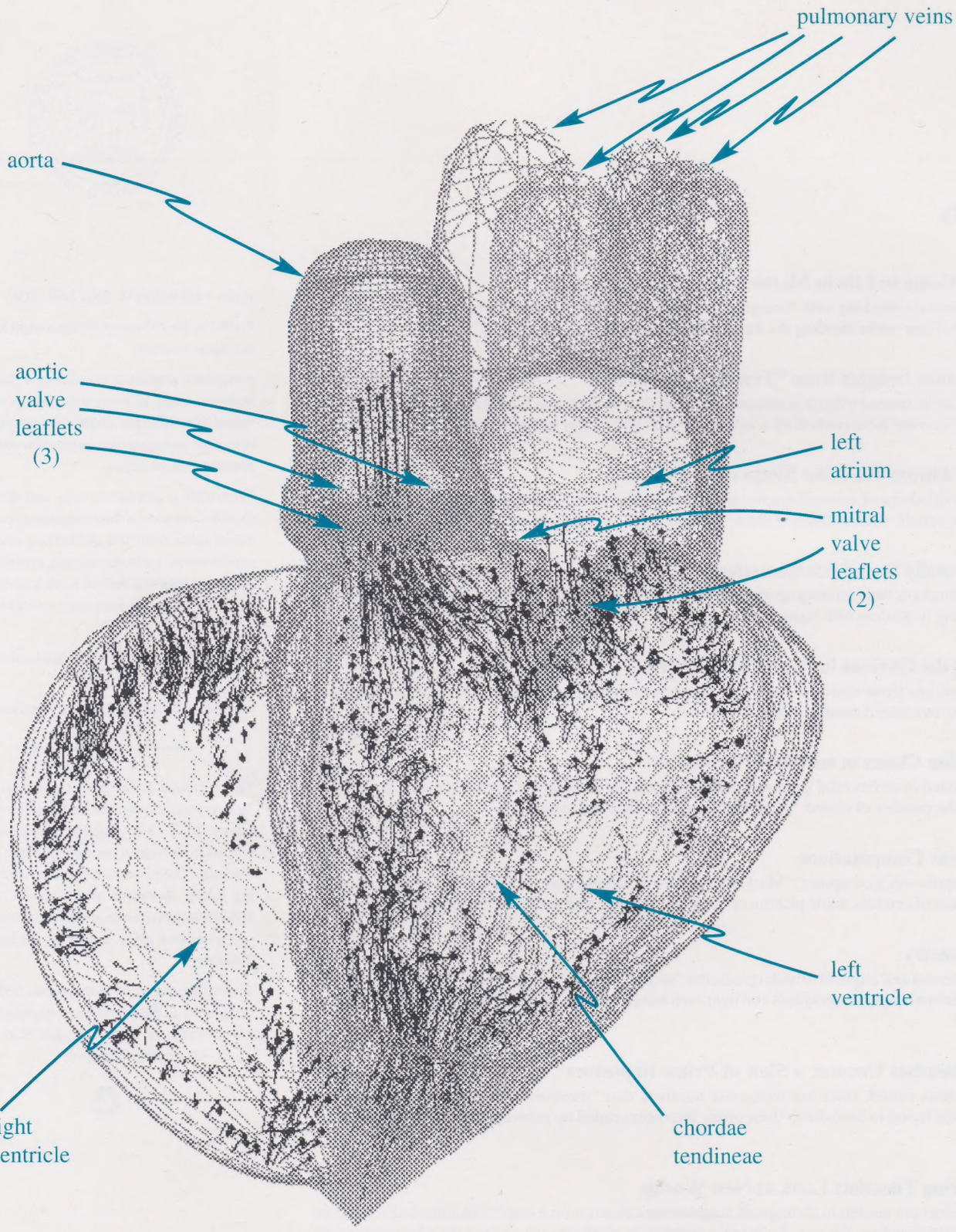
ISBN 0-8218-8999-0, ISSN 1065-9358

1991 *Mathematics Subject Classification*: Primary 00A06.

Printed in the United States of America.

This publication has been typeset using the TEX typesetting system running on a Solbourne 5/502 Unix computer. Halftones were created from original photographs with Adobe Photoshop and illustrations were redrawn using Adobe Illustrator on Macintosh Quadra and Macintosh IIci computers. PostScript code was generated using `dvips` by Radical Eye Software.

Typeset on an Agfa/Compugraphic 9600 laser imagesetter at the American Mathematical Society. Printed at E. A. Johnson, East Providence, RI, on recycled paper.



Researchers at the Courant Institute of Mathematical Sciences have created a three-dimensional model of the human heart; the computed flow pattern of blood is shown above. Grayish lines depict heart fibers, and black spots depict blood. The recent motion of the blood is indicated by the dark lines trailing behind the black spots. The figure shown is a single frame in the simulation of the blood flow, showing the flow pattern just after ventricular ejection. Figure courtesy of Charles Peskin and David McQueen.

Equations Come to Life in Mathematical Biology

Biology has a host of problems that call out for mathematical analysis, from the folding of proteins inside an individual cell to the complex food webs on the ocean floor.

The Nile crocodile and the Egyptian plover have a fascinating relationship. The croc, ordinarily a surly saurian, will sit placidly on the muddy river bank, mouth wide open, while the bird hops from tooth to tooth scarfing leeches and other tasty morsels. *Crocodylus niloticus* enjoys a thorough oral prophylaxis; *Pluvianus egyptius* gets a meal.

The technical term is *symbiosis*.

Something like that is evolving between biologists and mathematicians. Biology has a host of problems that call out for mathematical analysis, from the folding of proteins inside an individual cell to the complex food webs on the ocean floor. Mathematics, for its part, provides a quantitative framework that can bring order to the organic chaos of nature and point toward new directions for research. Mathematics has brought new insights into biology; biology has inspired new mathematical results.

Which you regard as the bird and which the crocodile is a matter, shall we say, of taste.

Mathematics and biology are not exactly newcomers to each other. Mathematical methods have long been used in population studies, epidemiology, genetics, and physiology, to name a few. And biological problems have spurred the creation of many mathematical techniques, including, arguably, the entire field of statistics.

What's new is the depth of detail that mathematical models are now striving for—and the attendant depth of theory required. The problems being tackled today call for closer cooperation than ever before between mathematics and biology. Increasingly, mathematicians are getting in on the ground floor of biological research, working directly with biologists to help tease out the mathematical structure in phenomena ranging from the undulating motion of fish to the beating of the human heart.

"The field's very different now than it was thirty years ago," says Alan Perelson, a mathematical biologist at Los Alamos National Laboratory and president of the Society for Mathematical Biology. "Early mathematical biology was really mathematics with a little inspiration from biology." There was little real communication between the fields. But the current generation of mathematical biologists, Perelson says, consists of researchers "who've been driven by the biology, who look at the details, talk to experimentalists, and generate models that are attempting to answer questions of interest to experimentalists."

Perelson's own work has been in theoretical immunology. He and others in the field are trying to develop mathematical models for the sequence of events that begins when, say, you step on a rusty nail, from the first antigenic signals presented by the invading bacteria, to the final chemotactic processes that close the wound either cleanly or with a lasting scar. It's not just a matter of programming a computer to do a bunch of calculations. Researchers first have to identify the crucial biological aspects of the process and then find the appropriate mathematical equations that describe them. Developing such a thorough understanding,

Nancy Kopell. (Photo by Janet Coleman.)

Perelson says, is the "grand goal" of theoretical immunology, but that goal is still a long way off. "We are really at the very beginning."

One reason for that is the daunting complexity of the immune system. The body's response to the variety of pathogens it encounters is, among other things, a pattern-recognition problem: The body must somehow identify an invading virus or bacteria based on the invader's distinctive pattern of chemical clues. The immune system's ability to do this, researchers believe, depends on the diversity of its receptors. "To do pattern recognition [for the immune system] seems to require on the order of ten million different types of receptors," Perelson explains. "So to understand in a profound sense how the system operates—to recognize pathogens and respond—one really has to deal with systems of enormous complexity." Mathematics enters the picture as a tool for modeling not only the individual receptors, but also the overarching structure that enables them to act in concert.

The emergence of organized behavior from a collection of individual entities is not unique to the immune system; it is a hallmark of living systems. A central problem in biology is to deduce how properties of a system at one level of organization produce behavior at higher levels—for example, how does the electrical activity in the nervous system of a centipede organize itself into the correct patterns to make the critter's legs move in a coordinated fashion?

Nancy Kopell, a mathematician at Boston University, likens this problem to the task of figuring out how a television works knowing only the properties of transistors. She sees the modeling of "emergent behavior" as a central concern for mathematical biology. "There are many questions in biology involving the behavior of systems in which what you can measure easily... is the behavior of some of the components of the system," Kopell says. "What you can't easily, or sometimes not at all, get from direct measurements is what's going to happen when you hook all these things up. For that you really need some kind of theory."

Kopell and her mathematical colleague Bard Ermentrout of the University of Pittsburgh have collaborated with biologists to study the rhythmic neuronal patterns that give rise to swimming in an eel-like fish called a lamprey. Researchers had known for some time that the electrical activity in the lamprey spinal cord could be represented mathematically as a "chain of oscillators"—something like a set of pendulums hooked together by springs, but with quite different mathematical properties. Kopell and Ermentrout formulated a new mathematical model based on a deeper analysis of how the oscillators are hooked together. Their model produced predictions which could be verified by experimentalists, and it provided new insight into how the electrical activity organizes itself to produce the swimming motion in lampreys. The model also helped point out new directions for biological research. And as new data from new experiments is found, Kopell and Ermentrout continue to refine their mathematics to better reflect the biology.

Computer simulation figures prominently in many of the modeling efforts in mathematical biology. Indeed, revolutions in both hardware and software have been crucial to advances across the board. The Human Genome Project, with its ambitious goal of mapping the roughly three billion base pairs that constitute human DNA, would be inconceivable without machines and mathematical algorithms for dealing with vast amounts of data. (It's not just a question of storing three billion pieces of information; it's a question of *analyzing* that data.) Likewise, mathematics is at the heart of much of medical imaging, including CAT scans, nuclear magnetic resonance, and positron emission tomography. These techniques

are made possible by machines that carry out mathematical manipulations of the data that pour into them.

One notable example of the use of mathematics and computer simulation in physiology is the work of Charles Peskin and colleagues at the Courant Institute of Mathematical Sciences at New York University. They are in the process of building a realistic three-dimensional mathematical model of the human heart. The model, they hope, will give researchers insight into the functioning—and malfunctioning—of real hearts and lead to improved designs for artificial valves and other replacement parts.

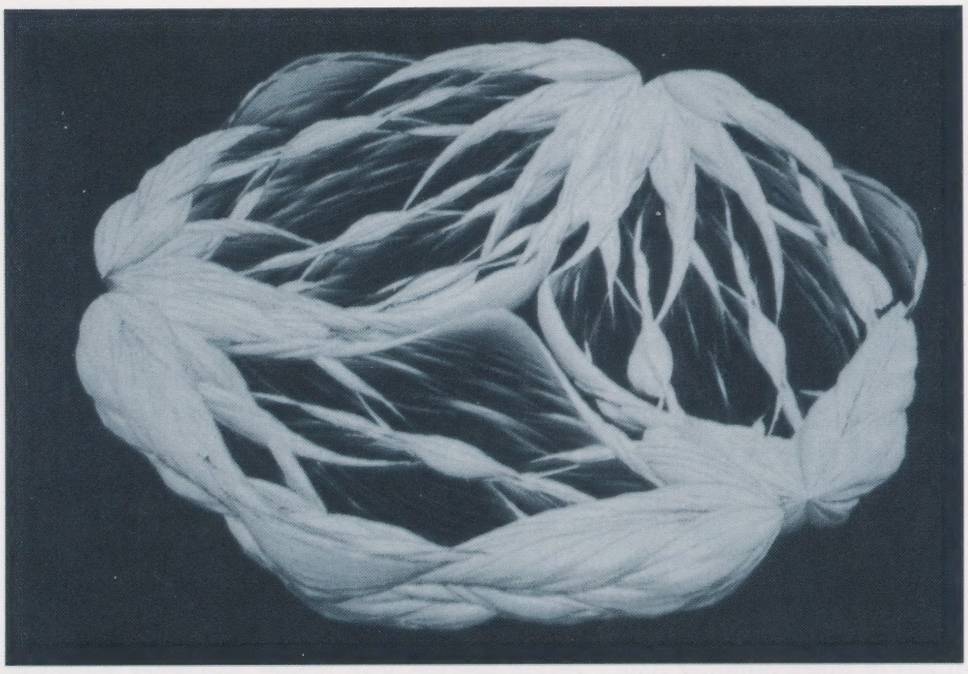
"It's a very large effort, and it's still going on," Peskin notes. The model is nearly complete anatomically, but "we're still working on getting the physiology right," he adds. That means figuring out the appropriate elasticities of the parts, how fast they should contract, and how fast they should relax, and then fine-tuning the equations to reflect these physiological attributes.

The geometry of the heart is also a crucial part of the model. Conceptually, the Courant heart consists of hundreds of closed curves representing muscle fibers. "In effect the [model] heart is constructed out of a very large array of rubber bands," Peskin explains. Mathematically, the curves are represented by a string of discrete points, with specified spring-like elasticity between each pair of consecutive points.

A computer keeps track of all these points—on the order of a million of them—and immerses them in a computer-simulated bath of blood. Then the real calculation begins: The numerical heart starts to beat.

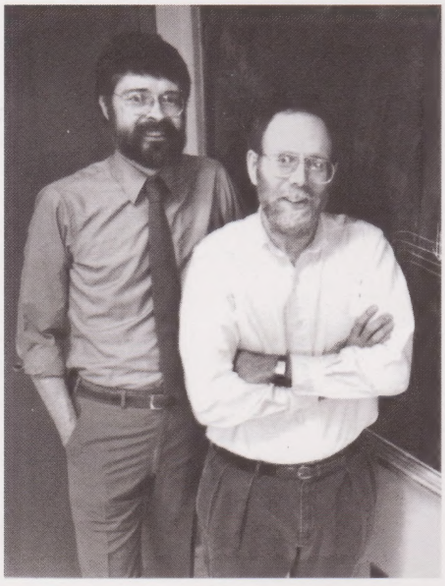
The mathematics of the calculation can be described by something that sounds like the title of a 1950s Japanese monster movie: Hookes's Law Meets the Navier-Stokes Equation. Hookes's Law is the force-displacement relation for springs, familiar from high-school physics; a fancier, nonlinear version of it is used to

Mathematics is at the heart of much of medical imaging, including CAT scans, nuclear magnetic resonance, and positron emission tomography.



Detail from the three-dimensional Courant heart, showing the three leaflets of the model aortic valve in its closed position. The fiber architecture of the valve has a fractal structure which has been predicted here by solving an equation for the mechanical equilibrium of the fibers under a pressure load. (Illustration created at the Pittsburgh Supercomputing Center.)

With the concurrent revolutions in both biology and applied and computational mathematics, [Peskin] says, "the kinds of problems that we can realistically hope to do are expanding tremendously."



David McQueen and Charles Peskin. (Photo reprinted with permission of Projects in Scientific Computing, *Pittsburgh Supercomputing Center.)*

model the heart's muscle fibers. The Navier-Stokes equation, while less familiar, is even more universal: It describes fluid flow of virtually any kind, from blood pumping through the heart to global circulation patterns of the earth's atmosphere.

These are the basic mathematical ingredients that determine the complex motions of the heart and the blood moving through it. Unfortunately, you can't sit down with pencil and paper and solve the equations precisely—the solutions are only approachable by computer approximation. And that turns out to be a formidable task, even for a supercomputer. Solving fluid flow problems is always computationally demanding, but the heart model presents a particular challenge: Unlike flow down a pipe or past a spinning turbine, where the boundary of the fluid is fixed or moving in a prescribed manner, the motion of the heart wall is among the unknowns that must be solved for.

"You not only don't know the boundary velocity, you don't know where the boundary is," notes David McQueen, a mechanical engineer who has collaborated with Peskin for the past fifteen years. "Your traditional engineering approach is going to be hard pressed to solve this problem."

Instead, Peskin has developed mathematical techniques for what he calls "immersed boundary" problems. "The beauty of this method is that it allows you to do computing in situations where you don't know the boundary motions in advance," says McQueen. Modeling heartbeats is not the only application. "The technique is generally useful in biofluid dynamics," Peskin says, "and it has already been applied to a wide variety of problems such as platelet aggregation during blood clotting, aquatic animal locomotion, and wave propagation in the inner ear." Peskin anticipates future applications in the study of flow in collapsible tubes such as thin-walled blood vessels, flow in renal (kidney) tubules, and the flight of birds and bats. There are even nonbiological possibilities, such as the design of aerodynamically efficient sails and parachutes.

The current heart model is a step up from a two-dimensional heart that Peskin began developing in the early 1970s. Paradoxically, Peskin notes, the 2-D heart is still, in some ways, more realistic than the 3-D model. That's mainly because the extra effort of computing in three dimensions has forced the modelers—for now—to use a simpler muscle model. Further advances in both theory and hardware will undoubtedly bring the 3-D model up to speed, but the 2-D heart is likely to continue being used for experimental computations. "What we'd really like is to use the 2-D model as a way of getting rough results, and then perhaps do a few 3-D computations to verify those findings," McQueen says.

Indeed, the 2-D model has already proved useful in artificial heart valve design. By experimenting with the shape of a prosthetic mitral valve (the gate between the left atrium and ventricle), McQueen and Peskin found a design that simultaneously increased the flow velocity near the valve and reduced the pressure drop across it—two features that are prized in artificial valves. While not yet in clinical use, the design has been patented and licensed.

The 3-D model has not yet had any such applications, but those are likely to come as the model becomes more physiologically realistic and as the computing demands get more manageable. (Currently a single beat takes upwards of fifty hours of supercomputer time.) Peskin sees the heart model, and other models in the future, as important experimental tools. With the concurrent revolutions in both biology and applied and computational mathematics, he says, "the kinds of problems that we can realistically hope to do are expanding tremendously."

New Computer Insights from "Transparent" Proofs

Mathematicians are professional skeptics. When told of a new result, their first response is, Where's the proof? Even when shown a proof, they're not completely convinced it's correct until they check every last line.

This professional skepticism isn't limited to traditional mathematical proofs. It extends to results produced by computers as well. Today's lightning-fast, high-tech adding machines take the labor out of long, laborious calculations, making it possible to carry out computations that could never be done by hand. But they leave behind the lingering question, Did the computer do its job correctly?

A sequence of recent breakthroughs in theoretical computer science may put that question to rest. Researchers have found some unexpected new ways by which computers can prove "beyond a shadow of a doubt" that the results they provide are indeed reliable. Moreover, these developments are giving theorists new insights into some of the hardest problems of computer science.

Guaranteeing the reliability of computer results is obviously of concern to more than mathematicians. But by thinking of computations themselves as proofs that certain inputs produce certain outputs, theoretical computer scientists are able to view anything a computer does in logical mathematical terms. Moreover, the computational aspects of many problems can be recast as purely mathematical questions in areas such as graph theory or elementary, first-order logic. The abstract language of mathematics helps clarify the essential issues, which might otherwise be lost among the details of individual applications.

Some computations are easy enough to check. For example, researchers often need to know if there is a path that travels along the edges of a graph, visiting each vertex once and only once—what graph theorists call a "Hamiltonian cycle." (This kind of problem crops up in applications such as designing efficient telecommunications networks.) If a computer says there is a Hamiltonian cycle, it can prove it simply by pointing out the path (as done with dark lines in Figure 1a). But when it says there is *no* such path for the graph in Figure 1b, how can you be sure it didn't overlook one—or, worse, that your computer saw one but chose not to tell you?

The computer can, of course, produce a proof by trying all possible routes around the graph and showing that none is a Hamiltonian cycle. That's not an unreasonable thing to do for Figure 1b. But the number of possible routes grows so quickly with the number of vertices that this straightforward approach soon becomes unwieldy. For graphs that typically occur in telecommunications network problems, for example, this kind of proof would take inconceivably long even on the fastest conceivable supercomputer. That defeats the purpose of having a fast machine. Worse, one is still left with the task of checking that all the computations were done correctly.

The problem is, errors in a proof don't always, or even usually, call attention to themselves—and all it takes to invalidate an entire proof is one mistake, as minor as a misplaced minus sign. "Mathematical proofs are very fragile," says László Babai, a theoretical computer scientist at the University of Chicago. Like a string

Today's lightning-fast, high-tech adding machines take the labor out of long, laborious calculations. But they leave behind the lingering question, Did the computer do its job correctly?

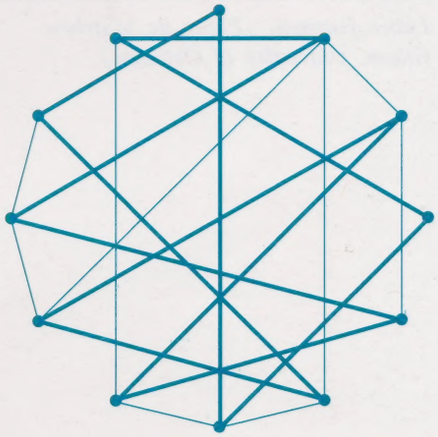


Figure 1a. *The dark edges "prove" that this graph has a Hamiltonian cycle.*

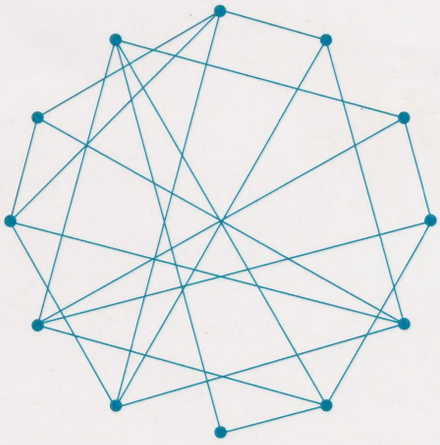
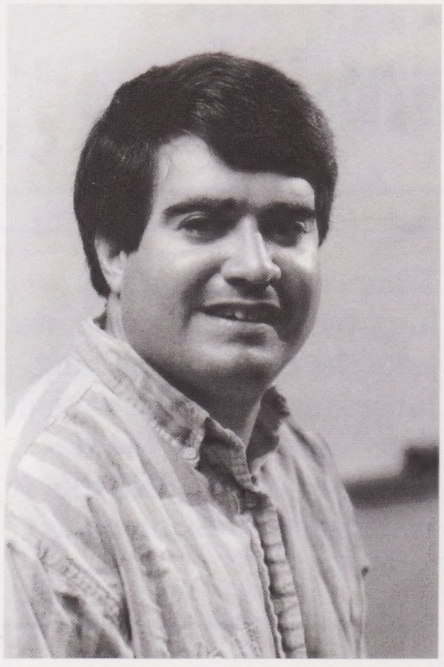


Figure 1b. *This figure does not have a Hamiltonian cycle—or does it?*



Lance Fortnow. (Photo by Matthew Gilson, University of Chicago.)

of pearls, if the strand breaks anywhere, the whole necklace winds up scattered on the floor—only with a proof, you might not notice till you've left the party.

However, help is on the way—sort of. Over the last decade, Babai and others have developed techniques by which even unreliable computers can, in principle, provide overwhelming evidence that their calculations are correct. Researchers have recently shown it's possible for a computer to reformulate an ordinary computational proof in such a way that the correctness of the original proof can be guaranteed—with near certainty—by merely "spot reading" the transformed version at a relatively small number of randomly chosen places.

That may not satisfy mathematicians, for whom being "nearly certain" is worth about as much as a basketball player's last-second, game-winning three-pointer that "nearly went in." The current techniques are also far from practical—it is unlikely the spot-checking techniques will ever be used directly to test the veracity of computers' output. However, the theory has paid off handsomely in other ways, mainly by giving researchers new insights in the theory of computational complexity—the study of how hard a computer has to work to arrive at an answer.

In particular, researchers have discovered an astounding connection with a seemingly unrelated issue in complexity theory: the question of whether there can be "easy" ways to approximate the solutions to computational problems in a class known as NP—problems that are thought to be intrinsically "hard" to solve exactly (see box on next page). Surprisingly, the existence of spot-checkable proofs turns out to preclude the existence of "easy" approximation algorithms for a substantial subset of the problems in NP—unless there are easy exact algorithms for the whole class NP, a prospect few in the computer science community believe to be the case.

The new results stem from work on "interactive proofs," a notion that was introduced in the mid-1980s by Shafi Goldwasser and Silvio Micali at MIT and Charles Rackoff at the University of Toronto. An interactive proof is a lot like a police interrogation. A "verifier" (the detective) asks a "prover" (the suspect)

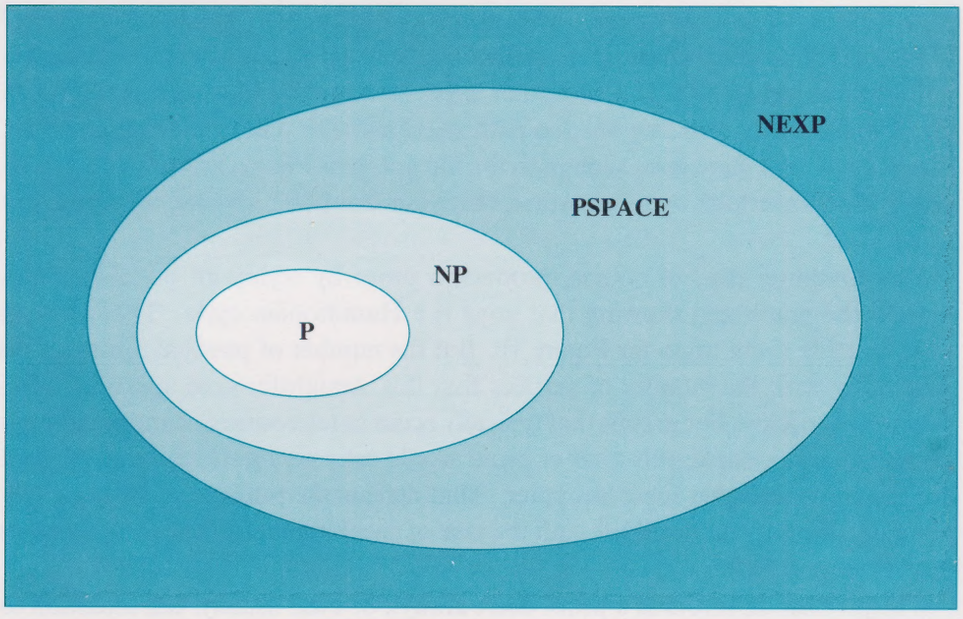


Figure 2. *Computational problems for which efficient algorithms exist constitute the smallest in a hierarchy of "complexity" classes.*

A Complexity Primer

The difference between "easy" and "hard" is at the heart of theoretical computer science. In essence, a computational problem is "easy" if the number of steps required to solve it is bounded by some power of the size of the problem. For example, multiplication of two N-digit numbers is "easy" because it takes at most N^2 single-digit multiplications and additions. Such problems form a class which computer scientists call P, for *polynomial time* problems (see Figure 2).

The class P contains a great many problems, including such significant computational tasks as linear programming. But a great many more problems seem to lie just beyond it, problems whose computational demands apparently grow exponentially with size. These are the problems that complexity theory calls "hard."

In particular, there's the tantalizing class of decision problems known as NP. (Decision problems are problems for which a simple Yes/No answer is sought. Technically P also consists strictly of decision problems, but when answers are easy to come by, the technicality is unimportant.) The problems in NP (which stands for *nondeterministic polynomial time*) have a curious dual nature: The amount of computation required to arrive at the Yes/No answer may grow exponentially with the size of the problem, but, at least when the answer is Yes, an "inspired guess" can reduce the amount of computation down to a simple, polynomial-time verification.

The Hamiltonian cycle problem is one example. The number of possible paths in a graph grows exponentially with the number of vertices and edges, but if a cycle exists, then all that computation is rendered unnecessary if someone simply tells you which path works, and you simply check it out. That can happen if the given graph was created by first drawing a Hamiltonian cycle and then disguising it with additional edges. In a sense, a problem in NP is a little like a riddle—only in NP, the riddle may have no answer.

But why, one may ask, is it necessary to check all possible paths in order to determine whether or not a graph has a Hamiltonian cycle? Is it not possible that some other method could arrive at the answer without going through an exponential number of cases? Is it not possible that the Hamiltonian cycle problem in fact is "easy"—a member of P—and it just looks hard because no one has found a polynomial-time algorithm for it yet?

Good question. In fact, that's the core conundrum of computer science. Nobody knows if there really are "hard" problems in the class NP; the classes P and NP may be one and the same. The lack of an easy algorithm for problems like the Hamiltonian cycle problem may be due not to the nonexistence of such an algorithm, but to the limits of mathematicians' ingenuity. It's within the realm of possibility that someone could find an easy algorithm for these hard problems.

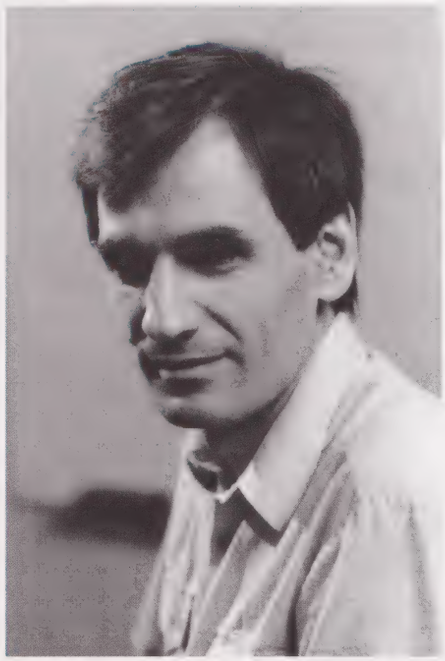
It's happened before. Linear programming was long classified as a hard problem because the simplex method was known to suffer the kind of exponential computational growth that's characteristic of NP problems. But then, in 1979, Leonid Khachian of the USSR Academy of Sciences discovered a polynomial-time—i.e., "easy"—algorithm for solving linear programming problems. Hence a problem that had been thought of as hard turned out not to be so hard after all.

It could happen again. But if it happens for the Hamiltonian cycle problem, or any of several thousand other NP problems, there will be a huge fallout. That's because over the last two decades researchers have established a web of relationships among the problems in NP. Specifically, there is a subclass of NP, the so-called "NP-complete" problems, which have the property that any algorithm for solving any one of them can be translated into an algorithm for solving any other problem in the entire class NP.

The Hamiltonian cycle problem is NP-complete. So is the well-known Traveling Salesman Problem. So are many other problems in graph theory, combinatorics, and logic. If anyone ever finds a polynomial-time algorithm for any one of these, the distinction between P and NP will collapse; only P will remain.

Most theorists consider that unlikely. Only a few die-hard optimists believe that all problems (at least the ones in NP) are easy; the smart money says that NP really does contain hard problems. But so far neither side is anywhere close to collecting on the bet.

The theory has paid off handsomely in other ways, mainly by giving researchers new insights in the theory of computational complexity—the study of how hard a computer has to work to arrive at an answer.



László Babai. (Photo by Matthew Gilson, University of Chicago.)

Transparent proofs are unaffected by minor copying errors or other computer glitches. In essence, a transparent proof replaces the original proof's single strand of logic with a highly redundant cable.

a series of questions about the problem the prover claims to have solved. The questions are designed to expose any lie (or mistake) in the prover's answers. In effect, interactive proofs are the embodiment of Walter Scott's familiar warning, "O, what a tangled web we weave, When first we practice to deceive."

In order to prevent a "mastermind" prover from anticipating the verifier's questions and concocting a consistent "alibi" to support its original claim, the questions are chosen partly at random. Because of this, there's a chance that an interactive proof will occasionally put its stamp of approval on an incorrect result, just as a lazy student can occasionally guess his or her way to a perfect score on a True/False test. But the chance of that happening can be made as small as you like by simply asking more questions.

Interactive proof turns out to be a powerful tool. In 1989, researchers established that interactive proofs can be used to verify solutions for a large class of problems called PSPACE. Then in early 1990, Babai, Lance Fortnow, and Carsten Lund at the University of Chicago proved what initially looked like an innocent generalization: They showed that problems in an even larger class called NEXP could be verified by a "multi-prover" variant of interactive proofs.

A multi-prover interactive proof can be thought of as an interrogation of *two* suspects who have been separated for questioning. The intuitive idea is that it's easier to get two suspects to contradict one another than it is to get a single suspect to trip over his or her own story. The Chicago theorists made that intuition precise. In demonstrating the exact power of multi-prover interactive proofs, they paved the way for what came next: "transparent" proofs.

The notion of transparent proof was introduced by Babai and Fortnow in joint work with Leonid Levin at Boston University and Mario Szegedy at the University of Chicago. In essence, they found that the question-and-answer format of an interactive proof is unnecessary; instead, all that's needed is to have the prover rewrite its proof as a kind of legal deposition—but one that's "easy to see through" if the prover tries to lie. This "transparent" proof is a long, rambling retelling of the original proof, couched in a kind of computer-science legalese, consisting of purportedly true statements which can be checked against each other for accuracy and consistency.

The key is that the correctness of a transparent proof can be checked without reading the whole proof, or even very much of it. Any error in the original proof, no matter how small, is magnified and spread throughout the transformed version so that it becomes glaringly obvious. By "spot checking" a relatively small number of randomly chosen passages of the transparent proof, the verifier—who can now be thought of as a judge—either finds a definite mistake or concludes, with very high confidence, that the original proof was correct.

This also means that transparent proofs are unaffected by minor copying errors or other computer glitches. In essence, the transparent proof replaces the original proof's single strand of logic with a highly redundant cable.

"You take a proof, which is fragile, and you turn it into a very sturdy thing," says Babai. In other words, if a transparent proof isn't riddled with errors, then the original proof is probably actually okay.

But how much spot checking is needed to be sure? Babai and coworkers showed that if the original proof was N bits long (remember, everything a computer does boils down to a string of ones and zeros), then the transparent proof could be written in such a way that the number of spot checks required to verify the

correctness of the original is proportional to a power of $\log N$, such as $100(\log N)^2$. That difference can be appreciated by comparing $N = 1{,}000{,}000$ to $\log N = 6$. What's most important is that any multiple of any power of $\log N$ is eventually an insignificant fraction of N, so for very long proofs the amount of spot checking will be relatively small.

This was taken a step further by Shmuel Safra at Stanford University and the IBM Research Center at Almaden and Sanjeev Arora, a graduate student at the University of California at Berkeley. Safra and Arora found a way to write transparent proofs that could be checked by looking at only about $\log \log N$ bits. Taken literally (using logs base 10), that implies that an original proof of length ten billion (10^{10}) could be checked by looking at a single bit of the transparent version!

But Safra and Arora weren't just out to reduce the spot-checking requirement of transparent proofs. They were after bigger game: an application of the new theory to an old and very important problem in computer science.

Shortly after the introduction of transparent proofs, Safra, Szegedy, and Goldwasser, together with Uri Feige and László Lovász at Princeton University, found an unexpected connection between interactive proofs and a particular problem in graph theory: that of approximating the largest "clique" in a graph of N vertices. A clique is simply a subset of vertices that are pairwise adjacent (meaning that there's an edge connecting each pair of vertices) (see Figure 3). The problem of determining the *exact* size of the largest clique in a graph is known to be NP-hard—that is, any efficient algorithm for solving this one problem would translate easily into efficient algorithms for solving any problem in the class NP.

What the five researchers showed was that the problem of *approximating* the size of the largest clique is "very nearly" NP-hard. In other words, if the size of the largest clique can be approximated—even poorly—by an efficient algorithm, then any problem in NP can be solved by algorithms that are "very nearly" efficient.

Safra and Arora removed those adverbs. Their refinement of transparent proofs implies that if the largest-clique problem could be solved approximately by an efficient algorithm, then there would be truly efficient algorithms for all problems in NP. In the jargon of computer science, NP would equal P.

That implication was soon extended from the largest-clique problem to a host of other approximation problems by Arora and fellow graduate student Madhu Sudan at Berkeley, Rajeev Motwani at Stanford, and Lund and Szegedy, both now at AT&T Bell Laboratories. They did so by pushing transparent proofs to an extreme: In their approach, all transparent proofs can be verified with the same number of spot checks no matter how long the original proofs are.

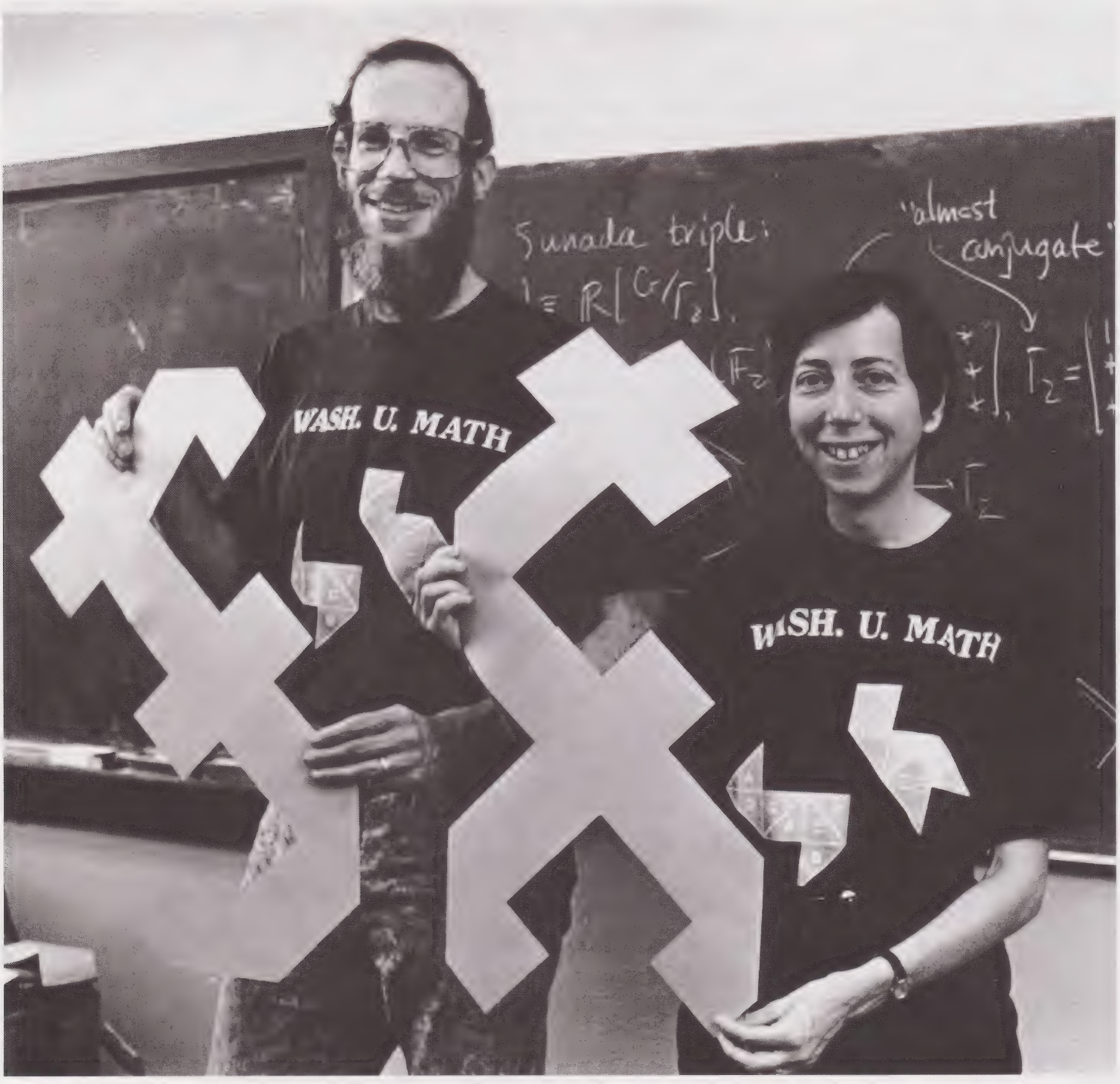
The only thing better would be a transparent proof you didn't have to read at all!

The string of breakthroughs in this area of computational complexity came in rapid succession—as befits a subject concerned with speed and efficiency. The implications for computer science—both theoretical and practical—are yet to be sorted out. Researchers want to know whether transparent proofs can be streamlined to more manageable lengths. They also are finding more problems that look "hard" to approximate. Finally, computer scientists continue to ponder what these results say about the class NP—in particular, are these seemingly hard problems really all that hard to solve?

The answer to *that* question is sure to have everyone looking carefully at the proof.

Figure 3. *The five vertices connected by the dark edges form a "clique" because there is an edge between each pair of points.*

The string of breakthroughs in this area of computational complexity came in rapid succession—as befits a subject concerned with speed and efficiency.



David Webb and Carolyn Gordon with paper models of a pair of "sound-alike" drums. (Photo courtesy of Washington University in St. Louis.)

You Can't Always Hear the Shape of a Drum

Much of what scientists know of the natural world comes not from direct observation, but by means of indirect measurements. Astronomers, for example, cannot sample the stuff of stars; instead they infer stars' composition by analyzing spectrographic images. Likewise, geophysicists construct a picture of the planet's interior from seismic studies, not from journeys to the center of the earth. X-rays, CAT scans, and other medical imaging techniques are also indirect ways of seeing inside the body. Even your family doctor prefers the stethoscope to the scalpel.

Mathematically, the job of reconstructing an object out of measurements of certain "observable" properties is known as an inverse problem. (A "direct" problem is to deduce observable properties from explicit knowledge of an object.) There are many important questions about inverse problems that mathematicians and others have worked to resolve, such as how many measurements are necessary to get an answer and how much accuracy is required. But underlying these questions is a deeper mathematical question: Even if you can take infinitely many measurements with infinite precision, can you be sure of your conclusions? Or to put it differently, can two different objects look alike in every measurable way?

It might seem the answer to this question should be obviously yes. But it's a lot more difficult than that—and that's where mathematical theory steps in. In 1966, the Polish-American mathematician Mark Kac zeroed in on a particular inverse problem. Can one, Kac queried, hear the shape of a drum?

That may seem like a strange question at first, but it's no stranger than asking if one can "see" the chemistry of a star or "hear" the interior of the earth. Moreover, Kac's question has a precise mathematical meaning. The problem it poses had been a challenge for more than fifty years at the time of Kac's lecture, and it continued to stymie researchers for another three decades. Then, finally, in the spring of 1991, three mathematicians—Carolyn Gordon and David Webb at Washington University in St. Louis, and Scott Wolpert at the University of Maryland—came up with the answer: a resounding No.

Gordon, Webb, and Wolpert found a pair of distinct geometric shapes in the plane which, when thought of as mathematical drums, resonate at the exact same frequencies. In other words, if your goal is to deduce the shape of a drum merely from the sounds it makes, these two drums provide an example where that goal cannot be achieved: You can't decide which drum you're listening to, because they both sound the same.

That's more than musically important, according to Dennis DeTurck of the University of Pennsylvania, an expert on "isospectral geometry," as the mathematical theory of such inverse problems is called. It points out there are subtle mathematical questions involved whenever scientists attempt to reconstruct reality from a set of data. The fact that even in a relatively simple mathematical setting there is not always just one conclusion that can be reached from a complete set of measurements is, to put it mildly, unsettling.

Remarkably, the final proof that the pair of sound-alike drums actually do sound

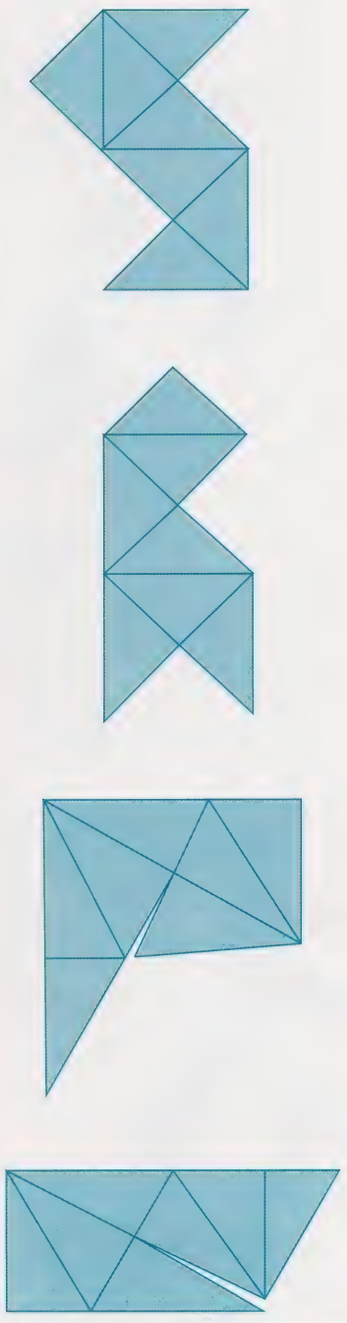


Figure 1. *The first two "drums" pictured above make the same "sound" although they are differently shaped. The same is true of the second pair.*

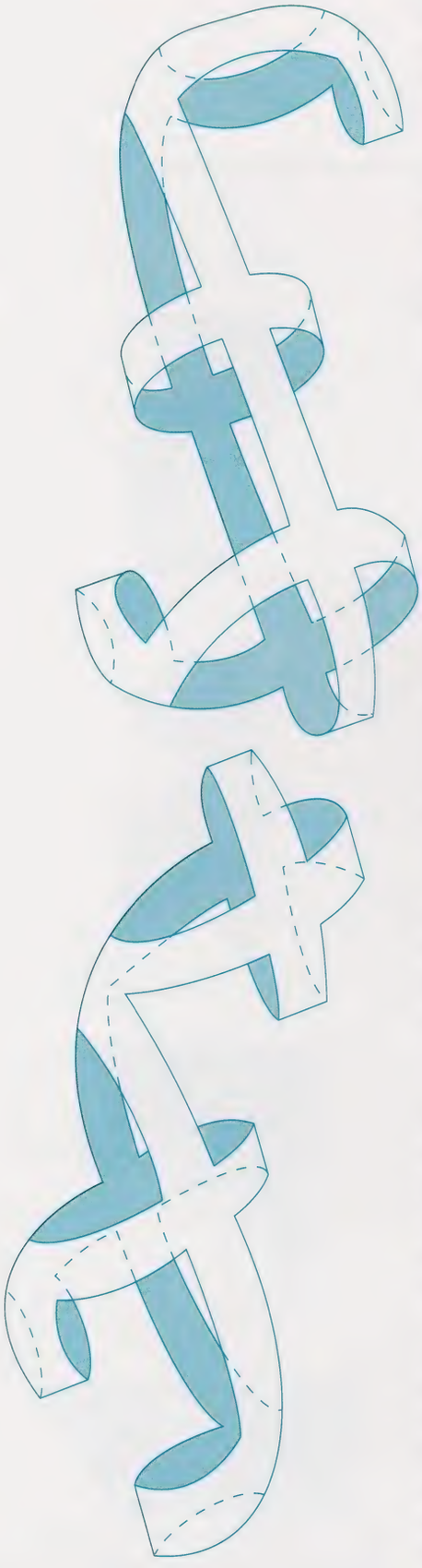


Figure 2. *A pair of Buser's isospectral "bells."*

alike involves little beyond elementary linear algebra. "It's amazing how simple you can make this proof," muses DeTurck. It fits "on a postcard," he says. In fact, the mathematics department at Washington University did something like that: They had T-shirts made up with the proof on them.

But while the proof itself is simple, finding the pair of drums to begin with was not. It required insights culled from careful study of geometric analysis, as well as new theoretical techniques that involve a surprising range of mathematical disciplines, from the theory of partial differential equations to representations of finite groups. At one point it also had Gordon and Webb (who are married) filling up their living room with huge paper models of geometric drums. And in the final stages, it had them running up a sizeable phone bill with transatlantic calls and twice-a-day faxes.

First, though, what is a "mathematical drum," and why should there be any connection at all between the shape of a drum and the sound it makes?

The first part is easy enough to answer. A mathematical drum is just a shape in the plane—a region with an interior and a boundary—such as a circle, a square, an arbitrary polygon, or just a blob surrounded by a smooth curve. The "sounds" produced by such a drum are determined by the solutions of a partial differential equation known as the wave equation, which is used to describe any kind of wavelike phenomenon, from sound to light to water. In essence, the motion of a vibrating membrane (that is, a drum) is governed by this equation, together with the condition that the drum not vibrate on its boundary.

That condition is crucial. Physically, it just says that the drum is attached firmly to a frame. Mathematically, it restricts the set of solutions to the wave equation. Without some sort of boundary condition, a mathematical drum could make any sort of sound.

Among the solutions to the wave equation are certain ones that are purely periodic in time—that is, vibrations that produce a single, clear tone of a specific frequency. While it's the interior that does the actual vibrating, it's the boundary that determines which frequencies are allowed. These frequencies constitute the sounds a given drum can make. They depend solely on the drum's shape.

Kac's question asked whether or not that dependence could be turned around.

There were reasons to think it might. In 1911, Hermann Weyl proved one can hear the *area* of a drum. Weyl's result accords with the intuition that the bigger the drum, the lower the tone. Some years later, the Swedish mathematician Ake Pleijel proved one can hear the length of the boundary. And Kac himself conjectured—and I. M. Singer and Henry McKean proved—that the number of "holes" in a drum is audible. These results made it plausible that the sound of a drum might contain enough geometric information to specify the shape uniquely.

On the other hand, there were good reasons to think that wasn't the case. In particular, mathematicians started finding counterexamples in higher-dimensional generalizations of the problem. John Milnor, now at the State University of New York at Stony Brook, found the first counterexample in 1964, a pair of geometrically distinct, sixteen-dimensional "isospectral manifolds"—that being the fancy term for "sound-alike drums." Over the next two decades, other researchers found additional counterexamples in lower dimensions. But these discoveries seemed to have no systematic basis. It was as if they arose by accident.

That changed in 1985. Toshikazu Sunada of Nagoya University introduced a method that made it possible to construct examples of isospectral manifolds almost at will. Sunada's method gave rise to a veritable cottage industry of low-

dimensional examples, including surfaces that can actually be cut out of paper and assembled with tape. These surfaces fail to answer Kac's question only because they aren't flat but rather curve around in three dimensions, more like bells than drums. However, it was one of these bell-like pairs, an example cooked up by Peter Buser at the École Polytechnique Fédérale in Lausanne, Switzerland, that ultimately led to the long-sought solution of the original problem (see Figure 2 on page 14).

The inspiration came at a geometry conference at Duke University in March of 1991. Gordon showed a paper model of Buser's bell-like example in a survey talk. Wolpert was in the audience.

"Scott came up to me after the talk and said he'd noticed that these paper models had a symmetry to them, and if you 'modded out' by a symmetry—meaning simply smashing them down—then you got plane domains," Gordon recalls. "So he asked whether they were isospectral. And that's what led to all this. It really is just smashing them down."

Wolpert's hunch was right. But it took a while to find the proof. The shapes that result from flattening Buser's example are too complicated to compute their sounds exactly, so a direct comparison was impossible. Moreover, Sunada's method did not apply to the kind of surfaces, called orbifolds, that were required to make sense out of the flattening process.

However, help was already at hand. Pierre Berard at the University of Grenoble had generalized Sunada's method to one that worked in the orbifold setting. He had also introduced a crucial notion of "transplanting" solutions of the wave equation from one manifold to the other in an isospectral pair. Berard's results were exactly what the Americans needed.

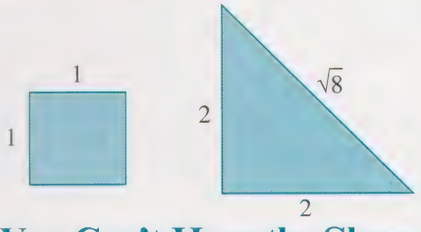
Even so, it took Gordon and Webb several weeks to find the right combination of ideas. They initially thought Buser's example was too simple to work, and spent days cutting out and taping together complicated paper models, looking for examples with good geometric properties. "We must have spent about a week, just building things out of paper," Gordon recalls. "And when we cut them in half, we must have spent about three hours trying to separate the two pieces [to get plane domains] before we realized they didn't separate!"

Finally they returned to the original example, which they had never completely abandoned. However, the last pieces of the proof came together while the two were thousands of miles apart—Webb at Dartmouth, Gordon in Germany. They hammered out the final details by phone and fax. By the time they got together, in Grenoble, they had a theory in place for a whole new class of orbifold-based isospectral manifolds, including a pair that lay flat in the plane, quietly answering No to Kac's old question.

That's not the end of the story, though. With help from Berard, Buser, and others, Gordon, Webb, and Wolpert identified the group-theoretic ideas that make the proof work, and now have a streamlined proof simple enough to fit on a T-shirt (see box on page 16). They also found other, simpler examples of sound-alike drums, some with as few as eight sides. Other researchers, including Peter Doyle and John Conway at Princeton University, have discovered additional shapes of elegant simplicity. (Using other methods, Conway and Neil Sloane at AT&T Bell Laboratories have found a family of four-dimensional examples similar to Milnor's original sixteen-dimensional example.)

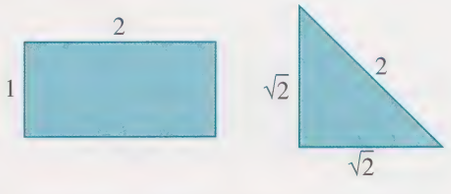
Gordon, Webb, and Wolpert's answer to Kac's question closes the book on one

Even if you can take infinitely many measurements with infinite precision, can you be sure of your conclusions? Or to put it differently, can two different objects look alike in every measurable way?



You Can't Hear the Shape of a Two-Piece Band.

Jon Chapman, a postdoc at Stanford University, took "scissors" to one of Gordon, Webb, and Wolpert's constructions and found a particularly simple pair of sound-alike drums, each consisting of two pieces. Chapman's two-piece drums, shown in the accompanying figures, are simple enough that it's possible to compute the exact sounds they make.



problem, but leaves others open—and raises new questions as well. "There are just tons of questions that come out of this answer," says an enthusiastic DeTurck. For instance, researchers right now only have examples of sound-alike pairs of drums; can there be sound-alike *triples*? Researchers also know that not all drums have isospectral twins (every circle, for example, makes its own unique sound); is there a way to tell which drums do and which ones don't? And do the group-theoretic techniques of Sunada's method provide a unified explanation of all isospectral plane domains, or are there other ways of constructing sound-alike drums?

Finally, questions remain as to what implications the negative answer to Kac's question has for other inverse problems and their myriad applications. The one thing that seems clear is that the techniques mathematicians have developed in studying drums are likely to have repercussions throughout the rest of science.

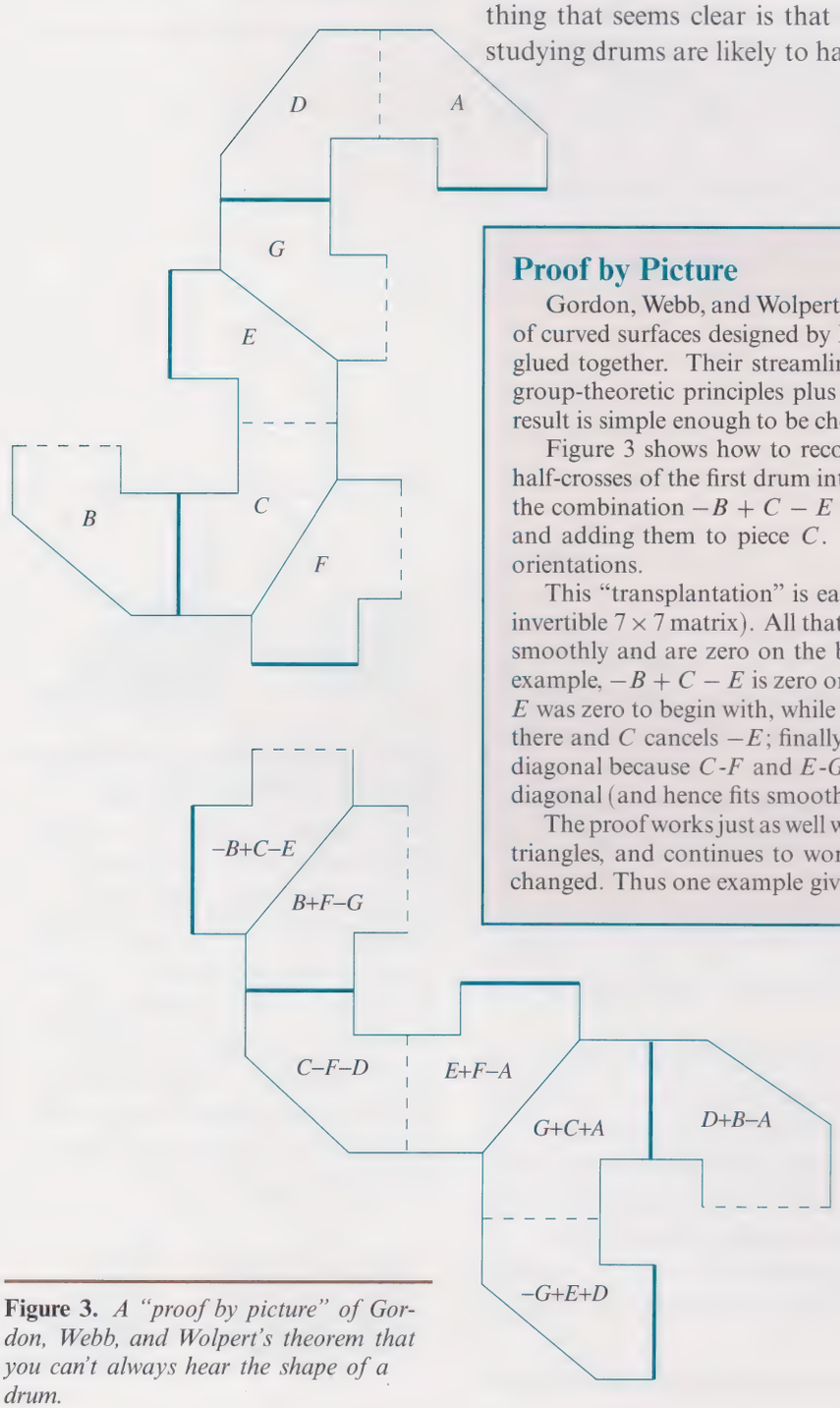


Figure 3. *A "proof by picture" of Gordon, Webb, and Wolpert's theorem that you can't always hear the shape of a drum.*

Proof by Picture

Gordon, Webb, and Wolpert's first example of sound-alike drums came from a pair of curved surfaces designed by Peter Buser. Each drum consists of seven half-crosses glued together. Their streamlined proof that the drums are isospectral is based on group-theoretic principles plus Pierre Berard's "transplantation" technique, but the result is simple enough to be checked directly.

Figure 3 shows how to recombine pieces A–G of a standing wave on the seven half-crosses of the first drum into a standing wave on the second drum. For example, the combination $-B + C - E$ is formed by "flipping" pieces B and E upside down and adding them to piece C. The dark and dashed lines emphasize the required orientations.

This "transplantation" is easily seen to work both ways (it can be written as an invertible 7×7 matrix). All that remains is to check that the combinations fit together smoothly and are zero on the boundary. But this can be done piece by piece. For example, $-B + C - E$ is zero on the dark boundary because $-B$ cancels C there and E was zero to begin with, while it vanishes on the dashed boundary because B is zero there and C cancels $-E$; finally, $-B + C - E$ fits smoothly with $B + F - G$ on the diagonal because C-F and E-G already fit smoothly together while B is zero on the diagonal (and hence fits smoothly with its reflection).

The proof works just as well when the half-crosses are shrunk down to right isosceles triangles, and continues to work if the angles of the triangles are (simultaneously) changed. Thus one example gives rise to an entire family of sound-alike drums.

Environmentally Sound Mathematics

Among the crucial scientific issues of our age, few are as far-reaching as those posed by the environment. Researchers from all fields have been called upon to investigate and evaluate the effects human activities are having upon the earth, from the upper reaches of the atmosphere to the depths of the ocean. The complex web of relationships in the biosphere demands an interdisciplinary approach.

Long the preserve of biologists, chemists, oceanographers, meteorologists, and geologists, environmental science is now drawing more and more upon the expertise of mathematicians as well.

Researchers in environmental science have long made use of mathematics to one extent or another. What's new is the recognition that rudimentary algebra and calculus are no longer enough to handle the sophisticated analyses that environmental scientists now know are necessary. Advanced techniques in scientific computing and numerical analysis are coming to the fore as researchers tackle challenging problems ranging from acid rain to the effects of the world's oceans on global climate.

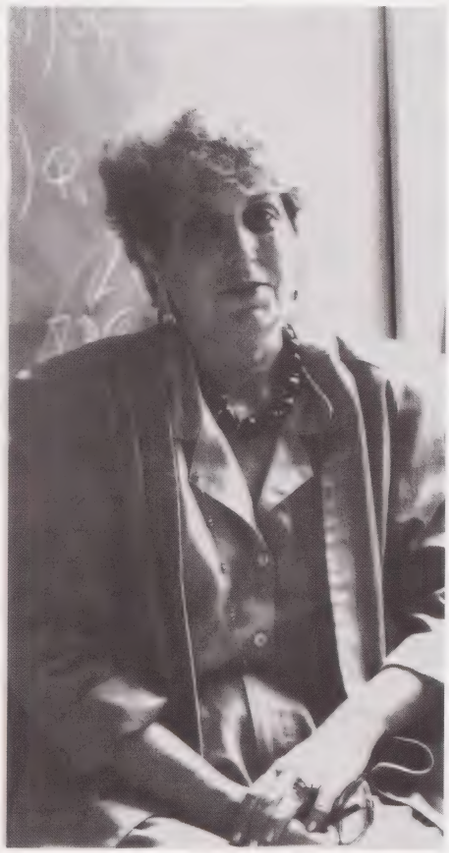
Rudimentary algebra and calculus are no longer enough to handle the sophisticated analyses that environmental scientists now know are necessary.

There is also growing interest in environmental science in the mathematics community itself. Recent meetings of the professional mathematics societies have featured presentations on environmental subjects, and last summer the Institute for Mathematics and its Applications, a mathematical think tank located at the University of Minnesota, held a four-week workshop on environmental modeling. Mathematicians are finding the field contains some interesting theoretical and computational problems. They are also finding a need in environmental studies for the ability of mathematics to build bridges between disciplines that are often separated by seas of jargon.

This kind of interdisciplinary work "takes time and energy, and it's not meant for everyone," says Mary Wheeler, a mathematician at Rice University, "but there are some really exciting challenges in it that will also drive some good results in mathematics."

Wheeler is one of the leaders in the movement of mathematicians into environmental science. She and colleagues at Rice and elsewhere have developed new mathematical tools for the study of fluid flow through porous materials. Their research combines the analysis of systems of nonlinear partial differential equations with sophisticated numerical algorithms that take advantage of new computer architectures such as massively parallel computation to solve the various equations. Among other applications, their efforts are aimed at helping environmental engineers plan remediation strategies for groundwater aquifers that have been contaminated by hazardous chemicals (see box on page 20).

One of the vexing aspects of environmental studies is the fact that the problems can span many scales of size. For example, a realistic climate model must consider everything from the microphysics of cloud nucleation to global circulation patterns. Likewise, plans for the isolation of nuclear waste must take into account physical processes occurring on a time scale of hours to months but keep an eye on safety standards valid for tens of thousands of years. In short, environmen-



Mary Wheeler. (Photo by Tommy Lavergne, Rice University.)

By working with mathematical models, environmental scientists can gain insights into systems that are too complex to study in any other way.

tal issues—take logging, for example—require researchers to look not just at the forest, but also at the trees (not to mention the spotted owls).

Mathematical modeling offers researchers the opportunity to identify and clarify mechanisms that connect phenomena at different scales, Wheeler says. In some cases sheer computational power makes the connection possible; in other cases, the mathematical equations themselves reveal the crucial interactions. By working with mathematical models, environmental scientists can gain insights into systems that are too complex to study in any other way.

Computers, and the high-tech algorithms that run on them, are making it possible to do the calculations required by these sophisticated mathematical models. "More and more people are recognizing that, with these tools, we can solve very complex problems," says Julius Chang, an atmospheric scientist at the State University of New York at Albany. Researchers no longer have to rely on unrealistic simplifications in order to make the computations tractable. "We can tackle many problems head on," Chang says.

Chang's group, for example, has developed an acid-rain model called RADM (for Regional Acid Deposition Model) which includes a system of coupled differential equations for a set of sixty different chemical species (see Figure 1). These aren't your nice, neat, textbook equations, either. Printed out, a typical RADM equation runs on for line after line of cryptic symbols and mixed upper- and lower-case letters, and could easily be mistaken for an old-fashioned computer core dump. RADM's equations take into account effects such as atmospheric advection and mixing, gas-phase chemical reactions, cloud mixing and "wet scavenging," dry deposition (acid can "fall" even when it isn't raining), and the location of sources of various pollutants (what comes down must have gone up).

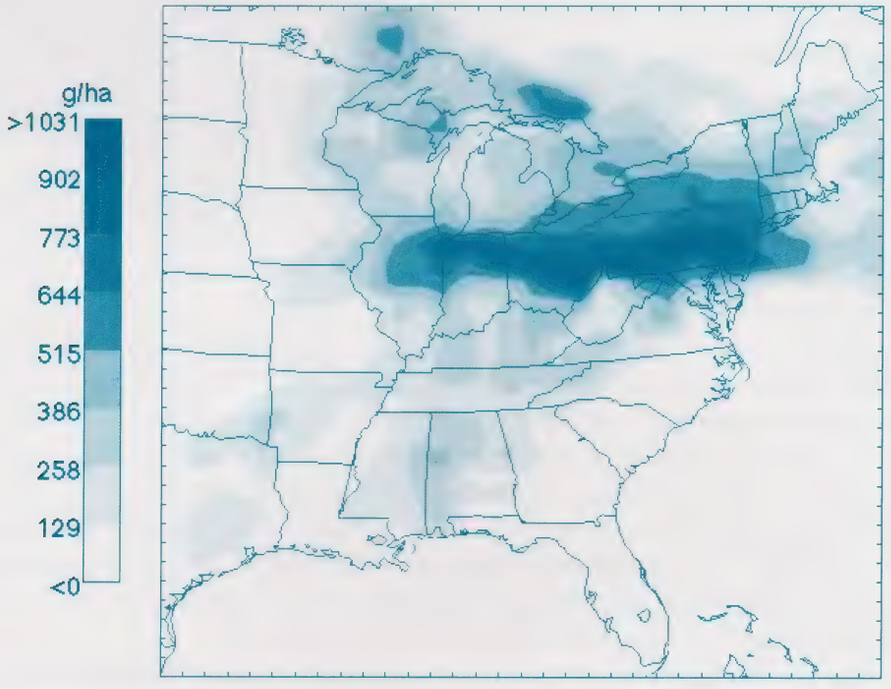


Figure 1. *Cumulative wet sulfur deposition (in grams/hectare) for eastern North America over a three-day period in April 1981 as calculated by the Regional Acid Deposition Model (RADM) using actual reported emission rates of various atmospheric pollutants. (Figure courtesy of Julius Chang.)*

Even so, there are limits to what computers can do. Take clouds, for example. "A cloud serves as an elevator for pollutants," Chang explains. But clouds are notoriously difficult to model. Scientists specializing in cloud processes have created "wonderfully detailed" models of cloud formation, Chang says, but their models "are too computationally complex to use on a regional scale. A single cloud can fill up a whole computer." So for now, RADM makes do with a cloud model that's fairly realistic but highly simplified.

If clouds are hard for computers to handle, oceans are even worse. There's a huge gulf between what goes on on the open seas and what even the largest supercomputer can model. That's a problem because oceans play a significant role in determining climate—and climate is one of environmental science's biggest concerns.

"If you're interested in climate, you're interested in the ocean," says Mac Hyman, a mathematician in the Theoretical Division at the Los Alamos National Laboratory in New Mexico.

The problem is not that oceans are hard to understand; in some respects, their equations are simpler than those of the atmosphere. Basically, the earth's oceans act as a gigantic heat reservoir and transport system. They exert a long-term influence on global climate by absorbing and emitting heat and carbon dioxide. The equations that describe all this are pretty well worked out. They include the Navier-Stokes equation, which underlies all fluid flow problems, and other partial differential equations describing gas exchange and heat transfer. All told, the basic equations of ocean dynamics can be written down on a single page. The problem is, these equations can't be accurately solved on existing computers—at least not by standard numerical methods.

That's because most of the kinetic energy of the ocean is found at scales too small for standard models to resolve, explains Hyman's colleague Darryl Holm. Unlike the atmosphere, where length scales on the order of 100 kilometers dominate the dynamics (just think of storm fronts), much of the ocean's energy exists in the form of eddies and waves that are five to ten times smaller.

"Without modeling some aspects of the small-scale, high-frequency waves and eddies, we can't know whether our global oceanic models are truly reliable," Holm says.

But researchers can't just tell the computer to take a closer look at the small-scale phenomena—or rather, says Hyman, "You can't afford the computer costs to resolve them. If you go down a factor of 10, that's a factor of 10 in three dimensions, so that's a factor of 1000—plus a factor of 10 in time, so that's a factor of 10^4, in computer costs, and no one's talking about those kinds of gains [in computer technology] in the next few years."

However, Holm and Roberto Camassa, also at Los Alamos, have developed some new mathematical approaches that may get around the problem. Their basic idea is to simplify the equations for ocean dynamics by taking advantage of the fact that certain important parameters, such as the ratio of surface wave amplitude to ocean depth and the ratio of depth to width of the ocean, are extremely small. If done carefully, the simplified equations will reliably represent the average effects of the high-frequency, small-scale elements on the large-scale dynamics. What should come out, says Hyman, is the correct average answer, "which is what we're looking for in the climate anyway."

Hyman, Wheeler, and others see a permanent role for mathematicians in en-

Mathematical modeling offers researchers the opportunity to identify and clarify mechanisms that connect phenomena at different scales, Wheeler says.

Mathematical modeling not only may help plan the cleanup of contaminants, Wheeler adds, it may also help contain costs.

vironmental science. After all, one of the things mathematicians do is to solve problems. And when it comes to the environment, the problems seem to be getting bigger all the time.

Bugs in the Program

The cleanup of underground aquifers contaminated by hazardous chemicals is serious—and costly—work. For example, the disposal of carbon tetrachloride over a period of eighteen years at the Hanford Site in south central Washington state has left contaminated groundwater over a five-square-kilometer area; cleaning it up could cost as much as $300 billion. And that's just one particularly bad example. The problem is not limited to a few locations. In 1986, the Environmental Protection Agency estimated there to be leaks in as many as 35% of the roughly 800,000 gasoline storage tanks in the U.S., with more than half reaching the water table.

"Contamination of aquifers by polluted streams and ponds, leaking storage tanks, agricultural chemicals, gasoline spills, and dumping has become a serious and widespread threat to public health," says Mary Wheeler. Wheeler has taken a keen interest in developing mathematical models that can assist environmental engineers plan their cleanup strategies.

One such strategy is known as *in situ* biorestoration. The basic idea is very simple: Certain microorganisms will actually digest or otherwise remove contaminants such as carbon tetrachloride—but only if encouraged to do so by the introduction of dissolved oxygen or other triggering nutrients. What makes it complicated are the complex interactions of groundwater, contaminant, organisms, and nutrients, which are flowing through material that may itself be highly heterogeneous.

That's where Wheeler's mathematics comes in. Wheeler and colleagues have developed mathematical models that describe these interactions in terms of nonlinear partial differential equations. They have also developed new computational techniques to solve these equations numerically and display the results using three-dimensional computer graphics. While noting there's still a lot of work to be done, Wheeler says these models should give researchers some much-needed insight into what's going on at contamination sites and how biorestoration can be used to best effect.

Mathematical modeling may not only help plan the cleanup of contaminants, Wheeler adds, it may also help contain costs. That's because "experiments" run on a computer are much cheaper than actual field experiments—and some experiments can only safely be tried in a computer "environment." The cost of conducting field experiments can run into the millions of dollars, Wheeler notes. "Doing it on the computer is very cheap."

Fighting over who gets to write down the next term in the equation: Darryl Holm (left), Roberto Camassa (center), and Mac Hyman. (Photo by Fred Rick, Los Alamos National Laboratory.)

Disproving the Obvious in Higher Dimensions

Not everything that's "obvious" is necessarily true.

Scientists in all disciplines know that drawing "obvious" conclusions, even from well-founded facts, is a dangerous game, unless those conclusions can be backed up by experimental verification. The same is true in mathematics, except that mathematical proof takes the place of laboratory experimentation. Mathematical explorations are guided by intuition, but only when their intuitions are confirmed by proof do mathematicians accept the "obvious" as true. This approach is necessary because sometimes what seems "obvious" just ain't so.

Mathematicians saw that happen not once, but twice in 1992. In similar but separate developments, researchers discovered that two facts from plane and solid geometry, facts that cry out for obvious generalization to geometric figures in *any* dimension, do not hold in that kind of generality. Their findings reaffirm researchers' suspicion that ordinary spatial intuition is not up to the task of thinking in higher dimensions.

That would be of only academic interest were it not for the fact that higher-dimensional geometry plays an important role in many mathematical applications. "I've been asked questions about higher-dimensional geometry by people who are interested in speech recognition and by people who are interested in algorithms for dealing with DNA," says Peter Shor, a research mathematician at AT&T Bell Laboratories in Murray Hill, New Jersey. Higher-dimensional geometry provides a natural mathematical framework for dealing with problems involving several variables or long strings of data. In particular, it has figured prominently in the development of so-called error-correcting codes, which are mathematical constructions that underlie the reliable storage and transmission of data in satellite telemetry, computer modems, and even compact disks.

One of the "obvious" generalizations was a problem that had been bothering mathematicians for the better part of sixty years before it was tackled by Jeff Kahn at Rutgers University in New Brunswick, New Jersey, and Gil Kalai at the Hebrew University in Israel. In 1933, the Polish mathematician Karl Borsuk proved that any region in the plane whose "diameter"—the largest distance between two points in the figure—is equal to 1 can be cut into three pieces, each of diameter strictly less than 1 (see Figures 1 and 2). This generalizes the completely trivial observation that a one-dimensional figure of diameter 1—that is, a line segment of length 1—can be cut into two shorter pieces.

On the basis of these two cases, Borsuk asked the obvious question: Is it always possible to cut any d-dimensional shape of diameter 1 into $d + 1$ pieces each of diameter less than 1? The "obvious" affirmative answer came to be called Borsuk's conjecture.

For many figures, of course, the task takes fewer than $d + 1$ pieces. The square of diagonal 1, for example, can be cut neatly in half. On the other hand, an equilateral triangle in the plane, a tetrahedron in space, and their cousins in higher dimensions definitely do require $d+1$ pieces: Since their vertices are all mutually a unit distance apart, each vertex must go into a separate piece. Borsuk's conjecture

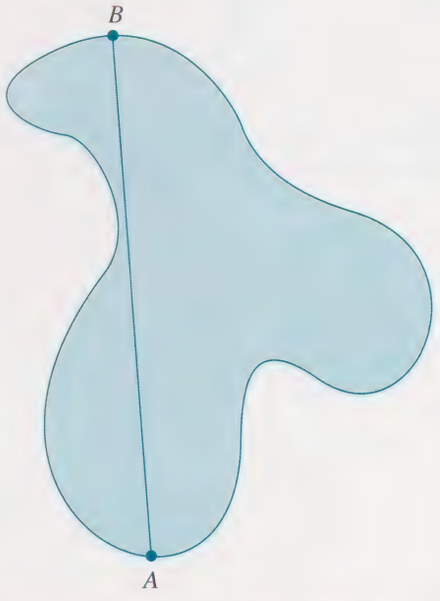


Figure 1. *A and B are the farthest apart of any two points in the shaded region. The distance between them is called the "diameter" of the region.*

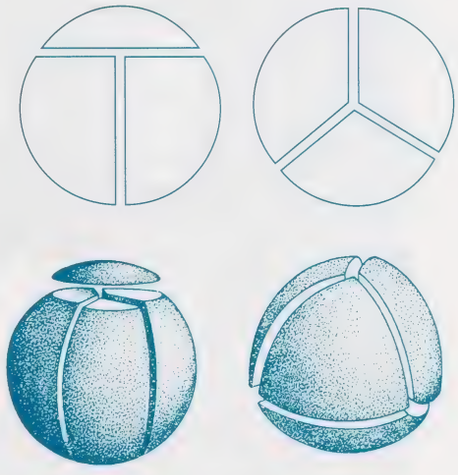


Figure 2. *Two ways to partition the circle and sphere into pieces of smaller diameter.*

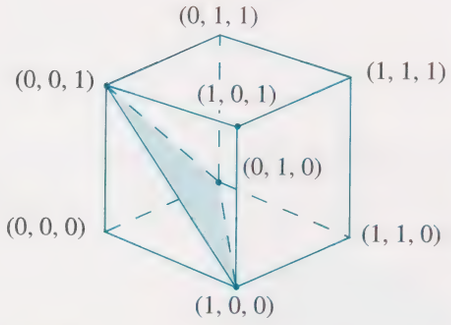


Figure 3. *A diagonal "slice" through the unit cube in three dimensions.*

One Plus One Equals 1.1?

If Borsuk's conjecture were true, then it should apply to any geometric figure. In particular it should be true for the vertices of a d-dimensional "slice" through the unit "cube" in $(d+1)$-dimensional space. The coordinates of the vertices of the unit cube are all zeros and ones, and one way to slice through it is to restrict to vertices that have a specified number of ones (see Figure 3).

Each vertex can be thought of as specifying a subset of the integers $\{1, 2, \ldots, d+1\}$ according to which coordinates are ones and which are zeros. For example, $(1, 1, 0, 0, 1)$ specifies the subset $\{1, 2, 5\}$. Under this interpretation, the distance between two vertices is related to the size of the intersection of the corresponding subsets: The smaller the intersection, the greater the distance.

In this setting, David Larman observed, Borsuk's conjecture reduces to a combinatorial assertion about sets: If S is a family of subsets of $\{1, 2, \ldots, d+1\}$ such that all the sets in S have the same number of elements and such that every two sets in S have at least n elements in common, then S can be partitioned into $d+1$ parts so that in each part every two sets have at least $n+1$ elements in common.

It was this version of Borsuk's conjecture that Kahn and Kalai found to be false. The vehicle they used to get there is a theorem of Frankl and Wilson: Let k be a power of a prime number, and let S be a family of subsets of $\{1, 2, \ldots, 4k\}$, each with $2k$ elements, such that no two members of S have k elements in common. Then S has at most $2\binom{4k-1}{k-1}$ members. (This bound is less, by an exponential factor, than the total number of sets with $2k$ elements.)

The reader is invited to ponder just how the Frankl-Wilson theorem contradicts Borsuk's conjecture. But remember Kalai's warning: "It's an example of an extremely short proof that was quite difficult to find."

doesn't say that *every* figure needs to be cut into $d+1$ pieces in order make all the pieces have smaller diameter; it just says $d+1$ is the *most* you ever need.

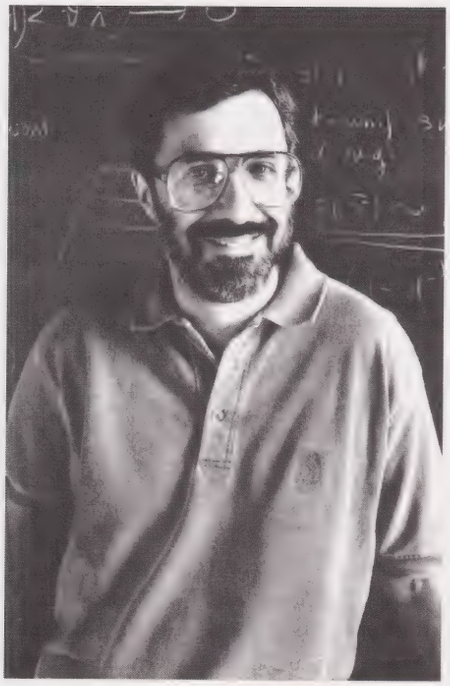
Things started looking good for Borsuk's conjecture in 1946, when the Swiss mathematician Hugo Hadwiger showed that any d-dimensional geometric figure can be cut into $d+1$ pieces of smaller diameter, if its boundary is smooth. In other words, Borsuk's conjecture is true for things like the d-dimensional sphere—shapes without corners or creases.

Then in 1955, the English mathematician H.G. Eggleston proved Borsuk's conjecture for $d = 3$. And that's pretty much where things stood until 1992, when Kahn and Kalai came in and knocked Borsuk's conjecture flat on its back.

Actually, Kahn and Kalai have not ruled out Borsuk's conjecture altogether; it might still be true in quite a few more dimensions. What they showed is that, for high dimensions, the minimum number of pieces required to cut any d-dimensional object into pieces of smaller diameter grows much more rapidly than $d+1$. Specifically, Kahn and Kalai proved that the minimum exceeds $1.1^{\sqrt{d}}$.

That formula doesn't do much good for small values of d, where $d+1$ is larger than $1.1^{\sqrt{d}}$. But starting around $d = 10{,}000$, the Kahn-Kalai bound kicks in. (The first instance where $1.1^{\sqrt{d}}$ is greater than $d+1$ occurs at $d = 9162$. With a little more care, Kahn says, they can obtain a formula that works down around 2000.) More to the point, their result shows that Borsuk's conjecture is *badly* wrong at high dimensions. The number of pieces needed grows exponentially, not linearly.

The form of the result might seem to suggest a long, complicated proof. After all, square roots don't often appear as exponents, and 1.1 is not the most natural number in the world. In fact, the proof is surprisingly short. It's only a few lines long. That doesn't mean the proof was easy to come by, though. "It's an example of an extremely short proof that was quite difficult to find," says Kalai.



Jeff Kahn. (Photo by Nick Romanenko, Rutgers University.)

Peter Shor and Jeff Lagarias. (Photo courtesy of AT&T Bell Labs.)

The proof is based on two ideas. The first, due to David Larman at University College in London, is an interpretation of Borsuk's conjecture as a statement about families of finite sets and their intersections. The second is a theorem due to Peter Frankl at the Centre National de la Recherche Scientifique in Paris and Richard Wilson at the California Institute of Technology about the size of such families (see box on preceding page). The hard part was "figuring out what to do with these ideas," Kahn recalls. Once they found the right construction, though, the contradiction to Borsuk's conjecture was an immediate consequence of the Frankl-Wilson theorem.

While it wipes out Borsuk's conjecture in general, Kahn and Kalai's construction of counterexamples leaves a lot of dimensions unaccounted for. In particular, "for dimension four, you clearly need a different way to look at the entire problem," says Kalai. The conjecture could be true or it could be false in that case. Nobody knows. And it could be another sixty years before anyone finds out. Or another six hundred years. Or it could be proved tomorrow.

Kahn and Kalai's cutting apart of Borsuk's conjecture was actually the second of the two counterintuitive geometric discoveries of 1992. Earlier in the year, Peter Shor and Jeff Lagarias, also at Bell Labs, took on another sixty-year-old problem, one with roots even older than that. The conjecture they looked at is based on a simple observation about squares in the plane: If you try to tile the plane with squares of equal size, then you necessarily wind up with squares that have an entire side in common. In fact there's essentially only one kind of tiling of the plane by squares, namely a checkerboard tiling in which the rows have been shifted by arbitrary amounts (see Figures 4a and 4b).

If you get your hands on a set of child's building blocks, you can convince yourself that something similar is true in three dimensions: If you "tile" space with cubes of equal size, you wind up with cubes that have an entire side in common. In this case, of course, the common side is a two-dimensional square.

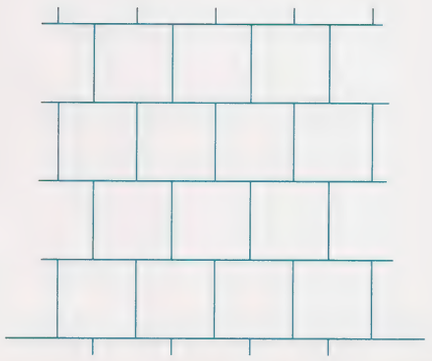


Figure 4a. *A tiling of the plane by unit squares.*

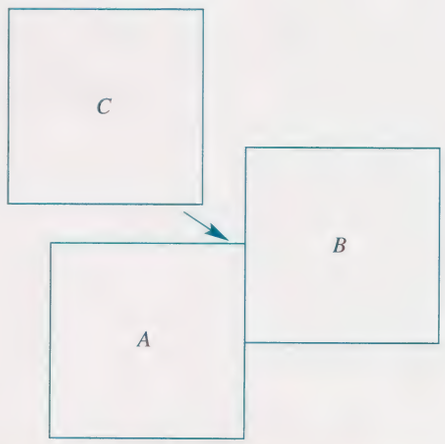


Figure 4b. *When square C is moved into the corner, it will have an entire side common with square A.*

What is clear, from both recent results, is that geometric intuition is a deceptive guide.

"Tiling space," whatever the dimension, means filling the entire space without any overlapping.

In 1930, the German mathematician Ott-Heinrich Keller took the plunge. He conjectured that no matter what the dimension d, if you tile d-dimensional space with d-dimensional "cubes" of equal size, then you wind up with cubes that have an entire $(d-1)$-dimensional "side" in common.

Actually, Keller was just generalizing a conjecture of Hermann Minkowski, who in 1907 made the same observation, but restricted to "lattice tilings"—tilings for which the centers of the cubes form a regular, grid-like lattice of points, like the locations of carbon atoms in a crystal of diamond. As it turns out, Minkowski was right and Keller was wrong.

Oskar Perron made his countrymen look good in 1940, proving Keller's conjecture in dimensions up to six. Two years later, the Hungarian mathematician György Hajós vindicated Minkowski completely, showing the original conjecture for lattice tilings is true in all dimensions. That left Keller's more general conjecture in dimensions seven and up. The issue lay unresolved for fifty years.

No more. Lagarias and Shor have found an explicit counterexample to Keller's conjecture in ten-dimensional space. This also kills the conjecture in dimensions eleven and up, because as soon as Keller's conjecture fails to be true at one dimension, it automatically stops being true at all higher dimensions. (A tiling by d-dimensional cubes can be converted into a layer of $(d+1)$-dimensional cubes, and then copies of the layer can be stacked to fill all of $(d+1)$-dimensional space, with the layers shifted so that there are no entire sides in common between layers.) The only unresolved cases are dimensions seven, eight, and nine.

In a sense, those cases require only patience—and maybe a high-speed computer the size of a major galaxy. That's because in addition to proving Keller's conjecture for dimensions up to six, Perron also showed how the conjecture could be checked in any given dimension by looking at a finite number of different tilings: If no counterexample is found in this finite set, then the conjecture is true (in that dimension). Unfortunately the number of tilings to be checked is unbelievably large: 2^{2^d}. No one in his right mind, and no mathematician either, would set out to sort through 2^{128} possible tilings to check the case $d = 7$.

Nevertheless, Lagarias and Shor did something of the sort to find their ten-dimensional counterexample. They based their construction on work of Kereszyély Corrádi at Eőtvős Loránd University and Sándor Szabó at the Technical University in Budapest, who two years earlier had introduced a new approach to looking for counterexamples. By studying the output of limited computer searches, Lagarias and Shor found tilings in dimensions three, four, and five that almost gave counterexamples in those dimensions. By cobbling these near-misses together, they manufactured legitimate counterexamples, first in dimension twelve, and then in dimension ten.

It's unclear if the same techniques can be brought to bear in dimensions seven, eight, and nine. Lagarias and Shor say it's possible the conjecture may fail even in dimension seven, but the counterexamples are too structureless to find. "The amazing thing is that there actually existed a counterexample that had a simple enough structure that you could actually find it," says Lagarias.

It's also unclear if their counterexample to Keller's conjecture will have any direct applications to things like error-correcting codes—so don't expect next year's line of CD players to be based on a tiling of ten-dimensional space. However, Lagarias

notes, the cube-tiling constructions give rise to novel types of "nonlinear" codes quite unlike the linear codes that are used in current applications.

What is clear, from both recent results, is that geometric intuition is a deceptive guide. "High-dimensional space is very strange," says Lagarias. Adds Shor: "If you're going to make conjectures about high dimensions, you should use some basis other than just extrapolation."

In fact, Shor goes so far as to make his own conjecture about higher-dimensional geometry: "Conjectures based solely on low-dimensional examples are false in high dimensions," he asserts. Asked if that includes his own conjecture, Shor amends the statement: "Conjectures based solely on low-dimensional examples are *likely* to be false."

Here's Looking at Euclid

The latest results involve some pretty highfalutin math, but not all counterintuitive results in higher-dimensional geometry are hard to prove. Here's one you can "see" for yourself.

Start by drawing four circles of radius 1 centered at the points $(1,1)$, $(1,-1)$, $(-1,1)$, and $(-1,-1)$ and then add a fifth circle centered at the origin and touching the other four (see Figure 5). This central circle is clearly contained in the square around the four outer circles.

The same thing is true in three dimensions: If eight spheres of radius 1 are centered at the points $(\pm 1, \pm 1, \pm 1)$, then a ninth, central sphere touching them all stays within the cube around the eight (see Figure 6).

It would seem obvious that no matter what the dimension, the central "sphere" always stays within the corresponding d-dimensional "cube." It's just not true.

Here's why. By the (generalized) Pythagorean theorem, the distance from the origin to any of the centers of the outer spheres is

$$\sqrt{(\pm 1)^2 + (\pm 1)^2 + \cdots + (\pm 1)^2} = \sqrt{d},$$

and consequently the radius of the central sphere is $\sqrt{d} - 1$. But the distance from the origin to any side of the cube is always just 2. So when $d = 9$, the central sphere touches each side of the cube, and for $d \geq 10$ it pokes outside the cube.

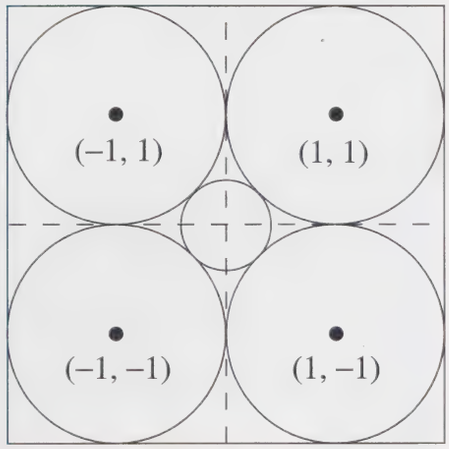


Figure 5. *The small inner circle touches all four circles of radius 1 and stays within the square.*

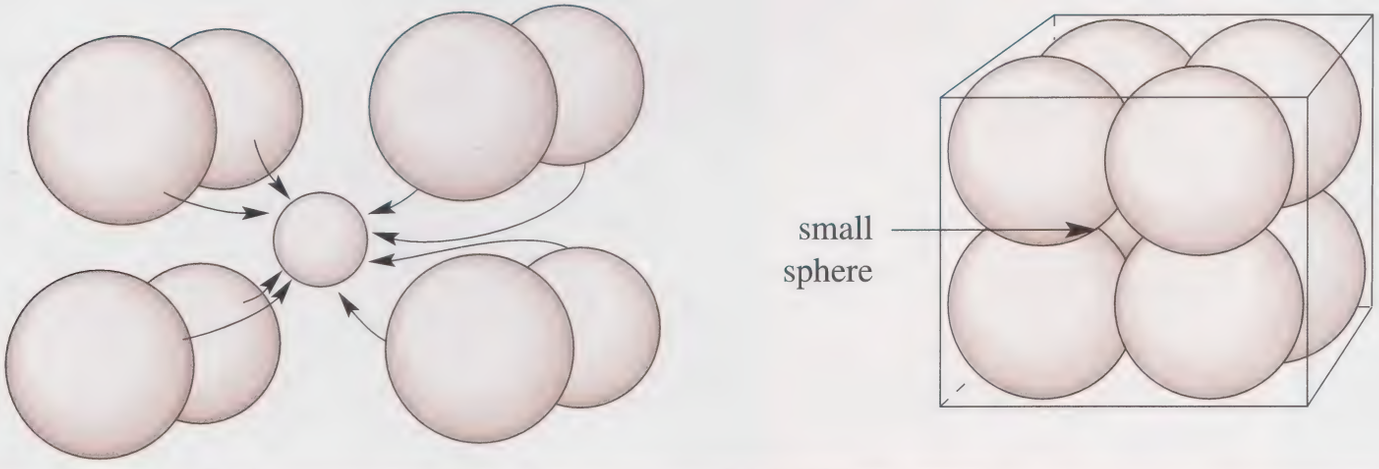


Figure 6. *Similarly, in three dimensions, the small inner sphere, which touches all the larger ones, remains inside the cube. This is no longer true in higher dimensions!*

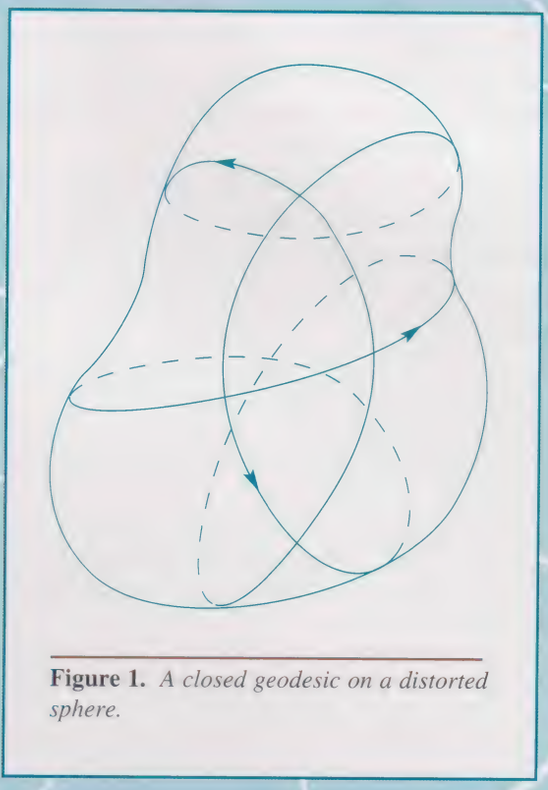


Figure 1. *A closed geodesic on a distorted sphere.*

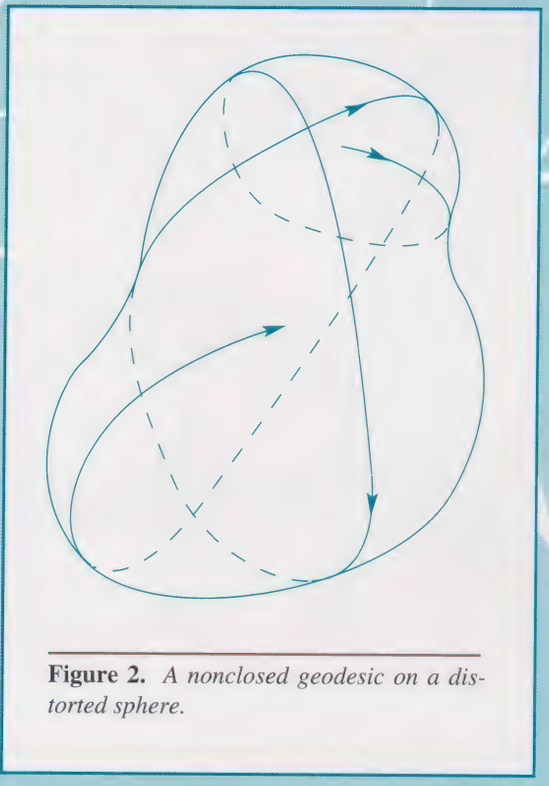


Figure 2. *A nonclosed geodesic on a distorted sphere.*

Collaboration Closes in on Closed Geodesics

Individually, neither hydrogen nor oxygen can combust. But put together and ignited by a spark, they are capable of exploding with enough power to propel a rocket into outer space. In mathematics, something similar can happen when two theories are brought in contact and set off by the spark of a new idea. Recently two mathematicians with expertise in separate specialties joined forces to solve a problem in differential geometry that had been on the books for more than sixty years.

Victor Bangert, at the University of Freiburg in Germany, and John Franks, at Northwestern University, have shown that no matter how badly you distort a sphere, there will always be infinitely many "closed geodesics" on it: rubber band-like curves that are determined by the curvature of the distorted surface (see Figure 1). The previous best result had been that every distorted sphere had at least *three* such geodesics—and that theorem dates back to the 1920s.

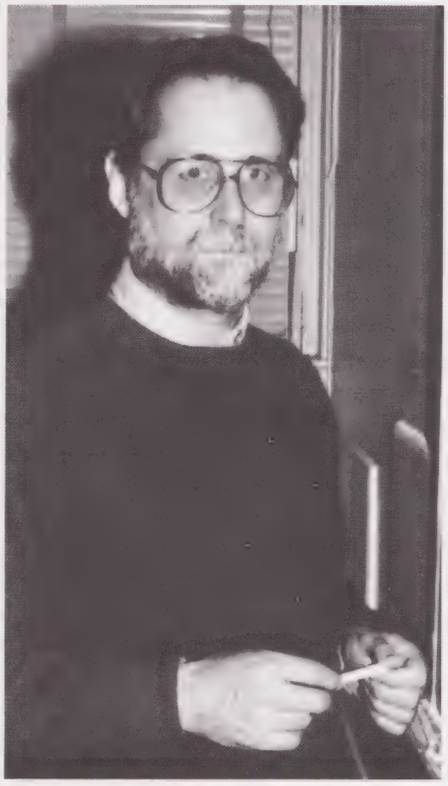
The new result is mainly of theoretical interest, but that doesn't mean it won't ever find practical applications. According to Robert Molzon, program director for geometric analysis in the Division of Mathematical Sciences at the National Science Foundation, differential geometry is applicable to "everything from general relativity and understanding the large-scale structure of the universe down to very small-scale problems such as boundaries between phases [e.g., liquid and gas] in materials science." Bangert and Franks's theorem is one more tool with which to study such problems.

Molzon is also encouraged by the new collaboration between two seemingly disparate mathematical areas: differential geometry and dynamical systems. "Bringing together these two areas is a big step," he says. Bangert and Franks solved the geodesic problem through a "divide and conquer" approach, with Bangert using classical techniques in differential geometry on one part of the problem and Franks bringing dynamical systems theory to bear on the other part.

Whether it's concerned with applications to relativity theory or materials science, or with more abstract issues in mathematics itself, differential geometry can be loosely described as the study of curvature. Geodesics are among its fundamental objects. A geodesic is basically just a path that follows the curvature of whatever surface or space it lies in. The precise definition implies that geodesics have a "shortest path" property. In particular, the shortest path between two points always lies along a geodesic.

The geodesics on a perfect (i.e., undistorted) sphere are the great circles, such as the equator or any line of longitude on the globe. Every one of them is closed. But as soon as you hammer on the sphere, that's no longer true. In general, when a geodesic traveling in one direction approaches a bump or a dent, it gets deflected in some other direction, much as a golf ball may veer away from the cup on an uneven green. It can easily happen that the geodesic will never find its way back to where it started (see Figure 2).

When there are bumps and dents everywhere, it's possible to imagine *every* geodesic wandering about endlessly. But that doesn't happen. George David Birkhoff proved in 1917 that every distorted sphere has at least one closed geodesic.



John Franks.

Whether it's concerned with relativity theory, materials science, or more abstract issues in mathematics itself, differential geometry can be loosely described as the study of curvature.

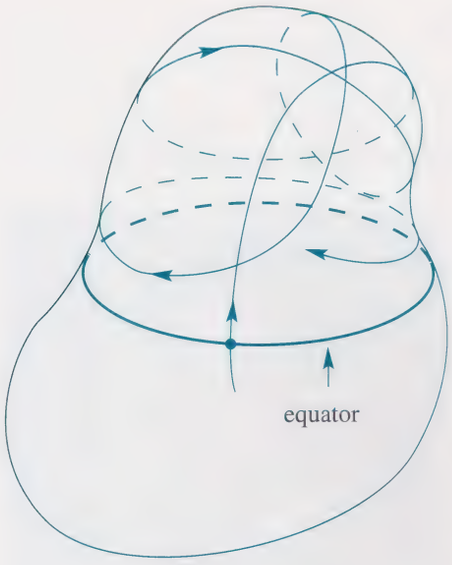


Figure 3. *A geodesic that gets "trapped" in the "Northern hemisphere."*

Twelve years later, two Russian mathematicians, Lazar A. Lyusternik and Lev Schnirelmann, went a step further. They proved there are always at least *three* closed geodesics on any distorted sphere.

More than half a century went by without a single closed geodesic being added to the count. Then Bangert had an idea for renewing Birkhoff's original attack on the problem. Part of the problem, Bangert saw, could be dealt with using techniques coming purely from differential geometry, his own specialty; the rest would require results from the theory of dynamical systems—and for that he sought the expert help of Franks.

"Bangert really kindled my interest in clarifying exactly what was needed to get this result," Franks recalls.

At first glance, differential geometry seems a far cry from the theory of dynamical systems. One subject is concerned mainly with objects that are fixed and permanent, while the other, virtually by definition, is interested in how things change. But the two aren't complete strangers. Birkhoff had already shown how the geodesic question could be translated into a problem purely in dynamical systems.

Birkhoff's translation starts with a single closed geodesic that loops once around the sphere without intersecting itself. This curve acts as a kind of "equator" separating two "hemispheres." Any other geodesic that crosses the equator either continues to cross it infinitely often (which is the case for closed geodesics if you keep following them around and around), or else it eventually gets trapped in one hemisphere (see Figure 3).

The former case, Birkhoff showed, leads to a dynamical system. Each crossing of the equator can be described by two parameters: one for the location of the crossing (i.e., its "longitude"), and one for the angle it makes with the equator. These parameters can be plotted on a washer-shaped region known as an annulus (see Figure 4). So each crossing of the equator by a geodesic corresponds to a point in the annulus, and, conversely, each point in the annulus corresponds to a crossing of the equator by some geodesic.

The theory of dynamical systems enters in when you follow geodesics from one crossing to the next. This defines a map of the annulus back onto itself—and maps from a region back to itself are one important kind of dynamical system. In particular, such maps can be iterated (that is, applied repeatedly). In the case of Birkhoff's annulus map, this corresponds to following a geodesic from one crossing to the next. The key point is that periodic points for Birkhoff's annulus map—that is, points on the annulus that eventually get mapped back onto themselves—correspond to closed geodesics. So to show there are infinitely many closed geodesics on a distorted sphere, it's enough to prove that Birkhoff's annulus map has infinitely many periodic points.

Bangert saw a division of labor. First of all, something had to be done in the case when Birkhoff's annulus map is not defined, which can happen, for instance, when a geodesic crosses the equator and gets trapped in the northern hemisphere. Bangert handled this case using classical techniques in differential geometry. In fact, his proof implies there are infinitely many closed geodesics anytime there are two geodesics that don't cross each other at all.

It remained to prove that, when the annulus map is defined, it's guaranteed to have infinitely many periodic points (corresponding to infinitely many closed geodesics). This case Bangert left to Franks, an expert on annulus maps.

Birkhoff's annulus map, it turns out, has a special property: When it maps the annulus back onto itself, it preserves the area, if not the shape, of any piece of the annulus. Birkhoff himself used this feature to prove that, at least under certain circumstances, his annulus map would have a fixed point, corresponding to a closed geodesic that intersects the "equator" in only one point. The proof, however, had nothing to do with geodesics or differential geometry; it was pure dynamical systems, a statement about area-preserving annulus maps.

Area-preserving annulus maps have been a staple of dynamical systems theory ever since. Franks had proved a generalization of Birkhoff's theorem (more properly called the Poincaré-Birkhoff theorem), and this was why Bangert approached Franks, in 1988, for help on the geodesic problem. It was clear what needed to be proved, Franks recalls. It just wasn't clear—at first—how to prove it.

Finally, in 1991, it became clear. Franks's theorem says that *any* area-preserving annulus map either has no periodic points, or else it has infinitely many of them. For the geodesic problem, the no-periodic-point possibility can be ruled out, and that leaves the long-sought conclusion: The sphere, no matter how badly distorted, still has an infinite family of closed geodesics.

Franks's theorem and Bangert's analysis don't completely close the book on the closed geodesic problem. If anything, the fact that there are always infinitely many closed geodesics raises a host of new questions. There are also questions raised by the proof itself. For example, among the closed geodesics on a distorted sphere, is there always one for which the Birkhoff map is defined? (If that's the case, then Franks's theorem alone would complete the proof that there are infinitely many closed geodesics.) The list of potential problems and new questions runs on—as endlessly as the geodesics themselves.

At first glance, differential geometry seems a far cry from the theory of dynamical systems. One subject is concerned mainly with objects that are fixed and permanent, while the other is interested in how things change.

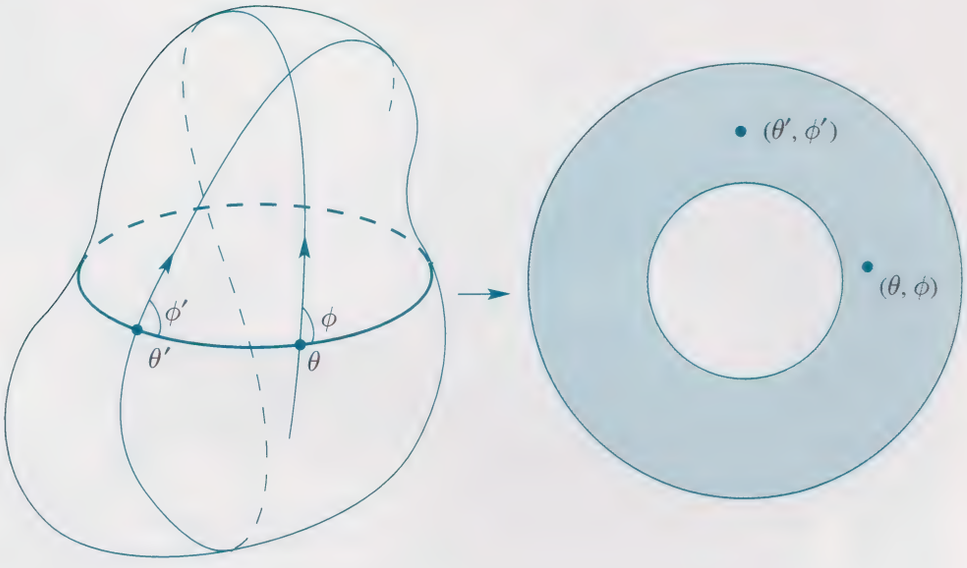


Figure 4. *The geodesic crosses the "equator" at a certain point along the equator (indicated by θ) and at a certain angle (indicated by ϕ). The values of θ and ϕ at any crossing are plotted onto the annulus.*

Rob Almgren and Andy Roosen at the Geometry Center in Minneapolis. (Photo by Barry Cipra.)

Crystal Clear Computations

Rob Almgren and Andy Roosen have a spirited competition going on. They're trying to see who can grow the nicest-looking crystals. Almgren and Roosen aren't working with chemicals in a laboratory, though. Instead, the crystals they grow are geometric shapes that develop in the purely numerical environment of a computer.

The two mathematicians—Almgren at the University of Chicago, Roosen a graduate student at Rutgers University—are members of a computational crystal growing group. The informal network of researchers, headed by Jean Taylor at Rutgers and Fred Almgren (Rob Almgren's father) at Princeton University, is part of a new trend in mathematics to combine the power of mathematical analysis with the speed and versatility of modern computers to tackle complex problems of fundamental importance head on.

The computational crystal growers are creating mathematical models and analytic techniques that will give scientists powerful new tools for studying the nature of crystals. These new tools, researchers say, will help accelerate the future design of materials with special properties of strength, "shape memory," and even superconductivity. At the same time, the computational crystal studies raise challenging problems in pure mathematics and numerical analysis, the solutions to which may well have applications in other, unrelated areas. The subject contains a "wealth of new geometric phenomena," says Taylor. "There are all these things out there waiting to be explained."

Snowflakes, for one. Just how the familiar six-sided crystal takes shape is still largely a mystery. Scientists know that the final shape depends on the conditions of temperature and supersaturation of water vapor while the snowflake is forming, but "the exact mechanisms are far from clear," according to Rob Almgren. The computational crystal growers will know they're on to something when they are able to mimic the growth of snowflakes.

But other applications are likely to come first. While snowflakes have intrinsic scientific (and aesthetic) appeal, there are also practical considerations driving research in crystal growth. The strength of steel, for example, is determined in part by the way crystals form as the metal cools from an initial, liquid state. Likewise, many properties of semiconductors depend on the way impurities in the melt are "driven out" in the process of solidification. Crystal growth is also crucial to the up-and-coming manufacturing technique known as molecular-beam epitaxy, in which materials are created one atomic layer at a time in a kind of ultra-high-tech version of spray painting.

Those phenomena all involve the growth of dendrites, and that's what Almgren and Roosen have been vying to recreate on their computer screens. Dendrites are structures that branch in complicated ways (the name comes from the Greek word for "tree"). In crystals, they are created by the interplay between surface energy and diffusion of heat or chemical impurities.

Surface energy in crystals is closely related to the area-minimizing surface tension that tends to keep soap bubbles and raindrops spherical. However, for crystalline materials the energy of a piece of surface depends on the direction it faces. Minimization of this "anisotropic" surface energy pulls the crystals into nonspherical shapes such as the cubical crystals of table salt. Heat and excess

Jean Taylor. (Photo by Rebecca Savoie.)

The computational crystal growers are creating mathematical models and analytical techniques that will give scientists powerful new tools for studying the nature of crystals.

The theoretical results "might never have been obtained without the new ideas generated by trying to do numerical computation," Rob Almgren notes.

chemical concentrations come from the release of latent heat and impurities whenever a portion of the liquid crystallizes.

"When heat is released, it must diffuse away," Roosen explains. Until it does, the region where it was released is too warm to crystallize further. "But what this means is that little bumps sticking out are able to diffuse their heat away faster than little dips, consequently they grow faster," he adds. "Without surface energy, this would happen with any bump or dip no matter how small, so the crystal would develop arbitrarily small structures. With surface energy, there is a lower limit on how small the fingers poking out into the cold can be." In other words, "release of latent heat and diffusion creates instability, surface energy controls it," Roosen says. The net result is the rapid advance of stable dendritic "tips" and the creation of characteristic branching patterns.

While that description seems straightforward enough, formulating it mathematically and then turning the equations into workable computer algorithms is a different matter entirely. Even in a simplified, two-dimensional setting, no one has yet come up with a method to match the rich range of structures seen in real experiments. Says Taylor: "Various people have announced that they 'understand' dendrites. I don't."

Almgren and Roosen's two-dimensional pictures look promising, however (see Figures 1a and 1b). Their methods often produce similar results, but they are based on different approaches. Both proceed by alternating steps in which the diffusion of heat is calculated with steps that compute the motion of the crystal surface. The main difference is in how they go about the second part of the calculation.

Almgren's approach treats the motion of the surface as a problem in geometric optimization. "At every step, you say 'What's the best shape that minimizes a certain energy function?' " he explains. Posing the problem in that way gives the approach an appealing conceptual generality. It also raises a number of theoretical questions and possibilities. In a good example of intergenerational as well as interdisciplinary research, Fred Almgren showed that the sequence of

Figure 1a. *A crystal grown using Rob Almgren's approach.*

optimizing shapes do indeed approximate a smoothly growing crystal. In fact, this established the existence of a solution to the original problem, which had not been known till then. The theoretical results "might never have been obtained without the new ideas generated by trying to do numerical computation," Rob Almgren notes. Refining the results "remains an area of active research," he adds.

Roosen uses the same algorithm as Almgren for computing heat release and diffusion, but his approach is otherwise quite different. For one thing, Roosen works with surface energy functions that are associated with "completely facetted" interfaces, meaning that the boundary of the crystal is a polygon with sides set at prescribed angles (Almgren works with smooth energy functions, for which the interface is always a smooth curve or surface). His approach is also "more direct" than Almgren's. "What I do is say, 'At this point, how does [the crystal] move?' And then I move it. What [Almgren] does is say, 'How does the whole thing move?' "

Roosen's crystals grow in a five-step process. The key step comes first: Each edge of the interface is moved according to a rule that depends on the temperature along the edge and the crystal's "weighted mean curvature" at the edge—a concept introduced by Taylor to make sense of curvature in a setting where the curves consist of straight line segments at prescribed angles (as dictated by the anisotropic surface energy). Taylor (who is Roosen's thesis advisor) has developed much of the theory that establishes motion by weighted mean curvature as a practical approach to computational crystal growth.

The second step in Roosen's algorithm is a merging process in which, for example, edges that have shrunk to zero length are removed from the program's bookkeeping system. Next comes a "shattering" step which takes into account the fact that some parts of an edge may actually want to move faster than other parts because of an uneven temperature distribution. In the final two steps, the program computes the release of latent heat and its diffusion. These five steps are repeated tens of thousands of times. A typical calculation, Roosen says, takes four to ten hours.

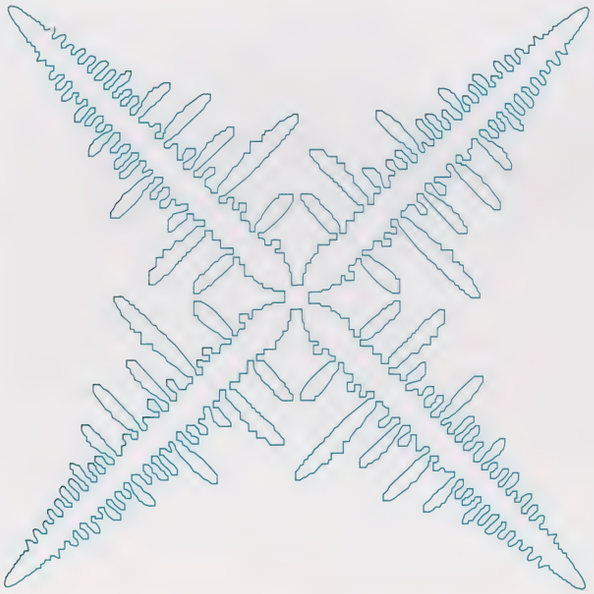


Figure 1b. *A crystal grown using Roosen's approach.*

That may seem like a long time to wait for a picture of a dendrite, but it's small potatoes compared to the run time for a full-blown three-dimensional computation. Both Almgren and Roosen's algorithms are conceptually suitable for three dimensions (much of the subject carries over easily into higher dimensions, for that matter), but the computational "load" can increase by a factor of several hundred. It would take weeks to simulate a single snowflake. It would take centuries to explore the variety produced in a single night's snowfall.

Improvements in the algorithms, further theoretical analysis, and more workstation horsepower are likely to bring 3-D calculations into the realm of practicality. However, another problem will remain: Figuring out a good way to visualize the results. Two-dimensional objects are easy to represent on paper or a computer screen; 3-D objects—especially objects you want to be able to see inside of—are far more challenging to represent.

Fortunately, the same issue crops up in a vast number of other problems, so a lot of thought has gone into this area. Researchers have made tremendous progress in recent years developing graphics programs that convert the computer's internal "knowledge" of an object into convincing, almost tangible pictures. It's something to look forward to: a 3-D movie (colorized, of course) of a computer-grown snowflake, surrounded by a glowing cloud of diffusing vapor.

The question is, will two such movies ever be alike?

A Growing Domain

Computational crystal growth is a wide-open field. Dendritic growth is only one of many open problems. Elizabeth Holm, a recent Ph.D. in materials science and scientific computation at the University of Michigan in Ann Arbor, is working on another: the structural evolution of "cellular arrays," such as occur in polycrystalline materials and—although it may seem far removed from the world of crystals—the foamy "head" on a glass of beer (see Figure 2).

As the name implies, polycrystalline materials are materials composed of many crystals, much as soap froth is composed of many individual bubbles. Over the course of time some domains grow while others shrink and disappear. The macroscopic properties of the material depend, in part, on the distribution of "grain" size in the crystalline microstructure.

Holm studies the process of domain growth with a computational model taken from statistical physics. The Potts model, as the approach is called, is like a "bitmap" of the microstructure, Holm explains. It describes the state of the material by an array of numerical indices assigned to a grid. In this setting, each domain consists of a contiguous set of grid sites that are assigned the same index. (On the computer screen, the indices are converted into colors.)

The evolution of the structure is modeled by any of a myriad set of rules. In one such rule, a single step of the algorithm is to pick a grid site at random, determine the number of neighboring grid sites with different indices, and if this number can be reduced by changing the index at the chosen site, then do so, otherwise either leave it alone or change it at random with some small, "temperature"-dependent probability.

Part of Holm's work has aimed at overcoming the effects of preferred directions (technically called "anisotropy") caused by the geometry of the grid. For example, in a square grid the boundaries between domains tend to be horizontal or vertical rather than diagonal. Holm and her coworkers have found that this inherent problem can be overcome in two ways: by extending the definition of "neighboring" grid site to a larger region (which, of course, entails more computation) or by increasing the "temperature" at which the simulation is performed. Their computations, including some 3-D simulations, indicate the Potts model should be useful for studying domain growth in a variety of physical systems.

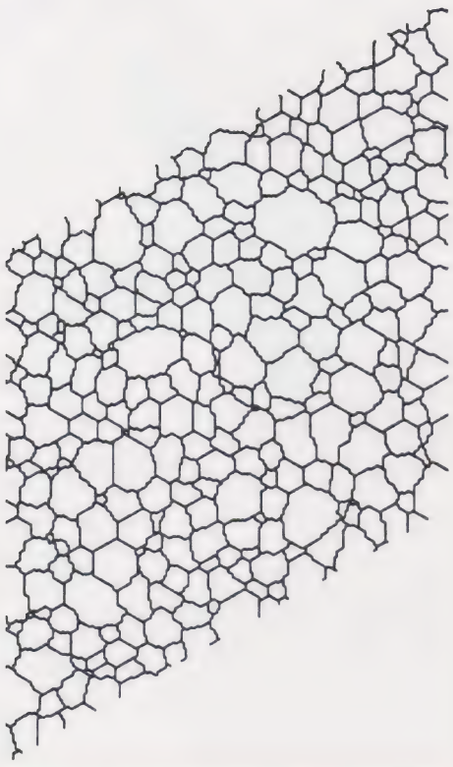


Figure 2. *Snapshot from Holm's computer study of the evolution of the grain microstructure in a two-phase polycrystal. (Figure courtesy of Elizabeth A. Holm, 1992.)*

Camp Geometry

You might call Kate Jenkins a budding mathematician. The Stanford sophomore spent last summer working on a computer program that parlays mathematical rules into pictures of flowering plants. Her geometrical bushes branch, bud, and even bend in a blowing breeze (see Figure 1).

Jenkins was one of nineteen undergraduates who participated in a summer research program at the Geometry Center in Minneapolis, Minnesota. The Geometry Center is a National Science Foundation (NSF) Science and Technology Center devoted to research at the cutting edge of geometry and computer visualization of geometric structures. But research is just one side of the coin; the Geometry Center also takes a serious interest in mathematics education.

Summer research programs for undergraduates in the mathematical sciences have become popular in recent years. The NSF last year awarded twenty grants in its Research Experience for Undergraduates program, at schools ranging from Williams College to Oregon State University. Many other colleges offered their own programs, as did research centers such as Los Alamos National Laboratory, the National Center for Atmospheric Research, and the Cornell National Supercomputer Facility.

"The philosophy is to give students a different experience with mathematics than the normal exam-packed classroom experience they get in school," says Al Marden, a professor at the University of Minnesota and director of the Geometry Center. The summer program gives students "a much more hands-on experience in mathematics, by doing it rather than by listening to somebody talking about it."

"It's sort of like an intellectual summer camp," adds Tony Phillips of the State University of New York at Stony Brook, who "coached" the students at the Geometry Center. For nine or ten weeks the students spent "all day and often part of the night" at the Center working on projects of their own choosing—"whatever they can think of," Phillips says.

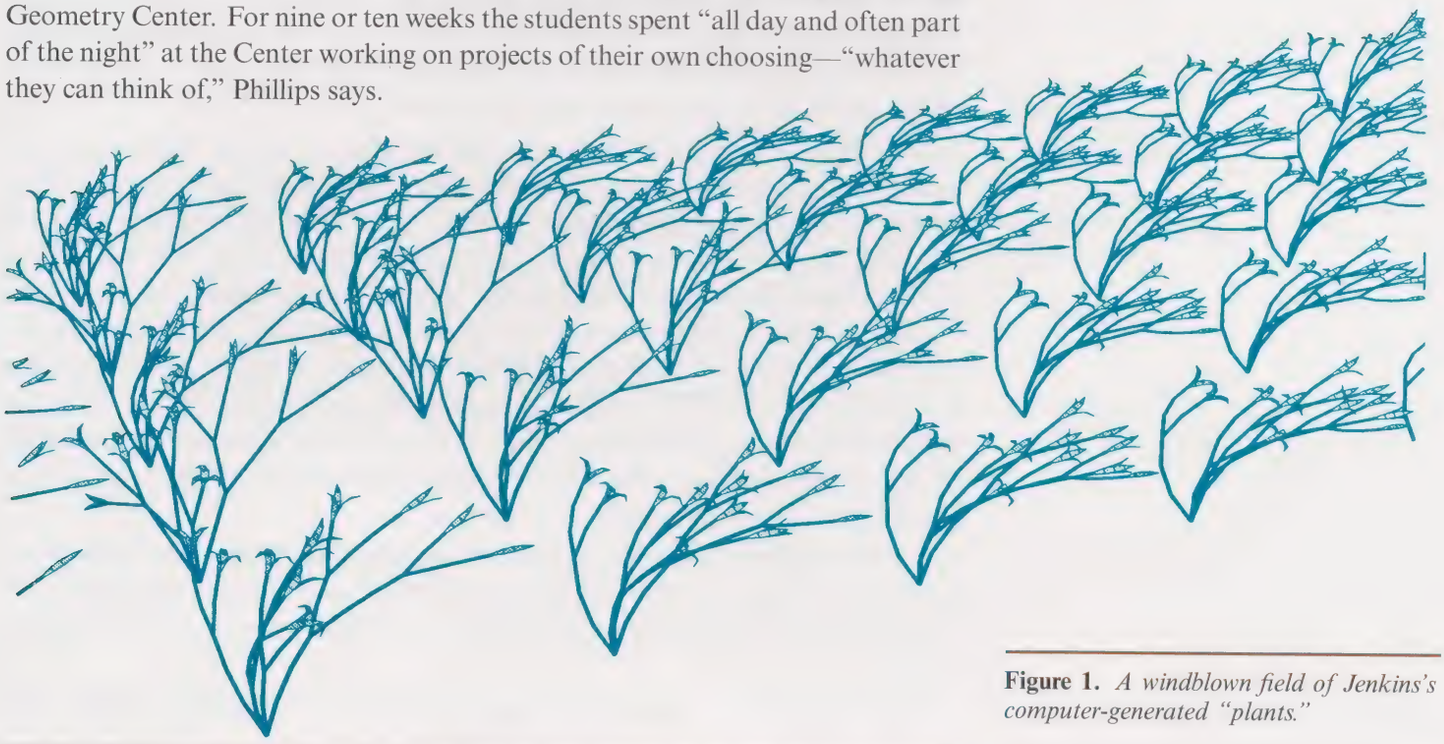


Figure 1. *A windblown field of Jenkins's computer-generated "plants."*

The summer program gives students "a much more hands-on experience in mathematics, by doing it rather than by listening to somebody talking about it," says Al Marden.

1991 Summer Institute participants. Center: *John Hubbard (Cornell University).* Front row, seated: *Stephanie Mason, Carol Sohn, Albert Marden (Director of the Geometry Center), Anthony Phillips (Head Coach, SUNY, Stony Brook), David Broman, Jennifer Ellison.* Second row, standing: *David Ben-Zvi, Mark Meloon, Adrian Mariano, Sherry Scott, Gary Gutman.* Back row, standing: *Jacques Friedman, Karen Olsson, Craig Sutton, Linus Upson, Nicholas Coult, Kate Jenkins, Ken Bromberg. (Institute participants not in photo: Chris Cianflone, Thomas Colthurst, and Prem Janardhan.) (Photo by Chris Faust, Space Science Graphics, University of Minnesota.)*

Jenkins took her cue from the book *The Algorithmic Beauty of Plants* by Przemyslaw Prusinkiewicz and Aristid Lindenmayer. She wrote a computer program that draws plants using "L-systems"—instruction sets that create complex forms by the recursive application of simple rules. L-systems were introduced in theoretical biology by Lindenmayer in the late 1960s.

L-systems give geometric life to an otherwise abstract algebra of symbolic manipulations. For example, an L-system might start with the character string FFRF, which it interprets as Move Forward, Move Forward, Turn Right, Move Forward. The key ingredient is a replacement rule which turns each instance of each character in the "instruction string" into some other string of instructions. In some systems, a single character— usually the "Forward" instruction— is replaced by the *entire* original instruction set. Thus, for example, FFRF becomes (FFRF)(FFRF)R(FFRF), or, removing parentheses, FFRFFFRFRFFRF. If the replacement rule is applied several times and then the resulting instruction string plotted by, say, drawing a line segment with each forward move, the result can be an elaborate, even organic-looking picture.

Jenkins employed more complicated branching and growth rules to produce animated "cartoons" of developing plants. Using a dash of vector geometry, she also worked out ways for her plants to rustle in a simulated breeze and dip due to gravity.

Stephanie Mason, a junior at Virginia Tech, also worked with L-systems, but

toward a totally different end: creating music. Mason takes the geometric result of an L-system and interprets it musically. For example, a vertical move may correspond to a step up or down in pitch, while horizontal moves indicate the duration of a note. The basic instruction set establishes a motif which iterations elaborate upon, she explains.

"You can actually create exactly what you want from these L-systems," Mason says.

One of her creations comes from an L-system that leads to a space-filling curve called the quadratic Gosper curve. Mason has set up the program so that the curve—really a set of line segments with right-angle turns—is drawn on a computer screen as the music is played on a synthesizer. Mason worked closely with Chris Cianflone, a student at the University of Minnesota (now in graduate school at the University of California at Berkeley), who developed an experimental musical program based on Fourier analysis of existing melodies.

Composers have long played with the formal structure of music. Bach, for example, is well known for writing music that could be played backward as well as forward. Mason has gone a step further, with music that can be played sideways as well, in what she calls a "right-angle canon." To do this, she simply takes a curve and rotates it so that pitch and duration are interchanged. When both curves are played together, using separate synthetic "voices" (Mason leans to piano and flute), the effect is surprisingly musical (see Figure 2).

"Bach would have loved it," Phillips remarks.

While Mason and Cianflone were turning Bach inside out, Nick Coult, a senior at Carleton College in Northfield, Minnesota, was putting a spring in orbit and numerically tracking the resulting motion. The idea, Coult says, was suggested by John Hubbard, a professor at Cornell who is on the permanent faculty at the Geometry Center. The problem is a variant on the three-body problem: There are

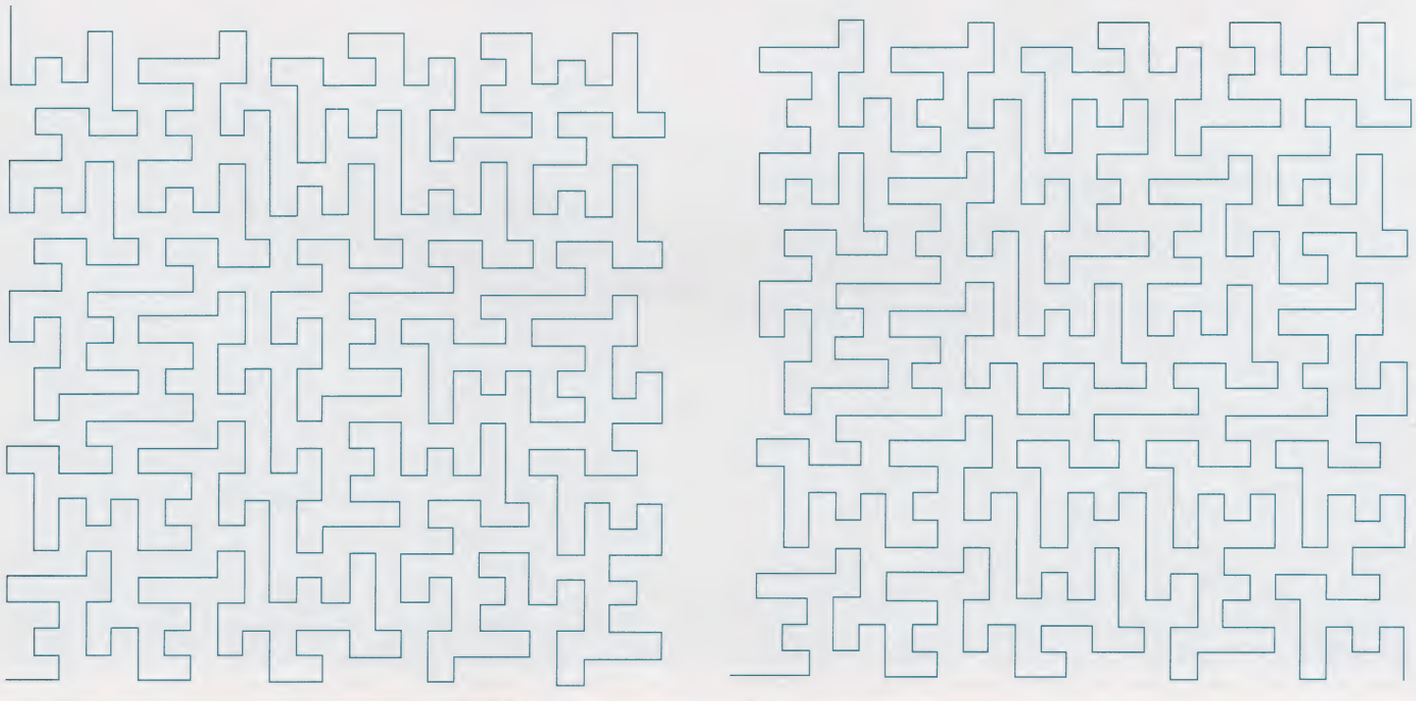


Figure 2. *Two examples of "musical scores" produced by Mason's computer program based on L-systems. The "flute" is on the left, the "piano" on the right. The flute starts in the lower left-hand corner, the piano in the lower right-hand corner.*

small masses at each end of the spring, and a large, gravitationally attracting mass which the mass-spring system orbits.

Hubbard had proposed the problem as a simplified model to test the hypothesis that tidal forces are responsible for the fact that we always see just one face of the moon. In this model, the earth is represented by the large mass and the moon is represented by the mass-spring system.

Coult indeed found that, in many cases, the system evolved into just such a stable orbit, with one end of the spring always pointing toward the central mass. However, he also found other stable orbits, some of which are surprisingly complex. "They're quite interesting to look at," Coult observes.

Setting up the equations that describe the motion was easy enough, Coult recalls. The challenge lay in solving them numerically, and then analyzing the solutions. For one thing, he says, "you have to verify that what you have on the computer screen [when it plots an orbit] is actually right." The programs he developed, Coult thinks, might be useful educational tools for courses in differential equations, although "that's not something I was thinking of when I started."

That sort of unexpected development is one of the benefits of letting students loose to do what they want in a relaxed atmosphere outside the usual classroom setting, Phillips points out. "Here, there's no test and there's no competition," he says. "The students are free to work at their own pace. They all work pretty hard, though."

Figure 3. *Another example of Jenkins's computer-generated "gardens."*

Number Theorists Uncover a Slew of Prime Impostors

Finding large primes and factoring their composite progeny have turned out to have applications in such areas as cryptography and computer security systems.

One of the oldest and best known examples of mathematical reasoning is Euclid's proof that the sequence of positive integers contains infinitely many primes. Three mathematicians at the University of Georgia have recently put a curious—and potentially important—twist on Euclid's famous theorem. W. R. "Red" Alford, Andrew Granville, and Carl Pomerance have shown that there are also infinitely many prime *impostors*.

No doubt that calls for an explanation.

A prime number, of course, is a number that's divisible only by itself and 1. Primes have lots of desirable properties; they are often called the building blocks of number theory. However, unlike building blocks, where it's easy to tell the difference between a block and a building, it's not always obvious whether a number—say 12345678910111213—is prime or composite (that is, a product of several primes).

At one time, the study of methods for identifying primes was considered an esoteric pursuit, even by many mathematicians. No more. Finding large primes and factoring their composite progeny have turned out to have applications in such areas as cryptography and computer security systems. The number-theoretic pursuit of efficient algorithms for primality testing and factorization also serves as a springboard for general ideas aimed at improving the efficiency of computer algorithms for other problems.

Computational efficiency is at the heart of the problem. If you're not worried about how long it takes, the definition of a prime gives you a simple way to tell if you've got one: Given a number like 12345678910111213, just try dividing it by 2, 3, 4, and on up to 12345678910111212. If one of these numbers divides 12345678910111213, then it's composite; if none does, then it's a prime. (Actually, it's only necessary to trial-divide up to the square root of the number in question, since the divisors of a number n can't all be greater than $\sqrt{n}$. Also, it's actually only necessary to trial-divide by *primes* up to the square root, since, for example, 6 can't divide a number if 2 and 3 didn't already divide it.)

Trial division works well when the number in question is small. But it's not a sensible way to verify the primality of large numbers. That's because the amount of computation it calls for gets quickly out of hand. For numbers with even just a few dozen digits, the computer run-times for trial-division primality testing start being measured in terms of the age of the universe.

Nobody wants to wait that long for an answer. It's like being put on hold when you're calling long distance.

But what can you do? Well, in 1640—long before the lightning-fast computers of today—the French mathematician Pierre de Fermat discovered a property of prime numbers that provides a surprisingly efficient test for primality—usually. What Fermat found is the following: If n is a prime number and a is *any* number whatsoever, then the number $a^n - a$ is divisible by n.

That statement has come to be called Fermat's Little Theorem (as distinct from the more famous "Last Theorem"). In spite of the diminutive title, Fermat's Little

The chance of being burned by a prime impostor is pretty low—legitimate primes are far more common than Carmichael numbers—but it's still worth reading the results with a *caveat calculator*.

Theorem is one of the most important results in number theory. In a very real sense it's the cornerstone of the subject.

For the purpose of primality testing, though, the value of Fermat's Little Theorem is in its ability to expose numbers that are *not* prime. It does this by turning the statement around: If n is a number and a is some number such that $a^n - a$ is *not* divisible by n, then n very definitely is *not* a prime number.

This turns out to be a remarkably efficient test: Given n, see if it divides $2^n - 2$, $3^n - 3$, $5^n - 5$ and maybe a few others (once again, it's only necessary to try prime numbers for a); if it fails *even once*, then n is not prime. For example, Fermat's Little Theorem "proves" that 6 is composite (if that were ever in doubt!)

Putting their heads together to solve a tough number theory problem: Andrew Granville (top), Red Alford (left), and Carl Pomerance. (Photo by Rick O'Quinn, University of Georgia News Bureau.)

because 6 fails to divide $2^6 - 2 = 62$. The fact that 6 does divide $3^6 - 3 = 726$ makes no difference; for a number to be prime, it must *always* pass the test imposed by Fermat's Little Theorem.

What makes this a good test for primality is that composite numbers tend to be exposed very quickly. Most of the time it's only necessary to test $2^n - 2$—the smallest composite number that slips by that test is $n = 341 = 11 \times 31$. It's very rare for a composite number to pass Fermat's test for more than a couple of values of a.

But it does happen. In fact, there are composite numbers that pass Fermat's test for *all* values of a. The smallest such number is $561 = 3 \times 11 \times 17$. Then come 1105, 1729, 2465, and 2821. These numbers are impostors; they "masquerade as prime numbers," says Granville. He calls them "annoying."

Infinitely annoying.

The first examples of these prime impostors were found around 1910 by American mathematician Robert D. Carmichael, and for that reason they are called "Carmichael numbers." It would be one thing if there were just a few of them—that would offer hope that Fermat's Little Theorem could be used not just as a proof of compositeness, but also as a guarantee of primality. But theorists kept finding more of them, and it seemed likely the list would prove endless.

That's exactly what Alford, Granville, and Pomerance have now shown: There are infinitely many Carmichael numbers. Moreover, their work makes it clear that many, if not all, other primality tests based on ideas similar to Fermat's Little Theorem are equally flawed by infinite families of composite numbers that pass the various tests. "There's no way to just generalize Fermat's Little Theorem to a [perfectly accurate] primality test," says Alford—or rather, he adds, "there's probably no simple way" to do it.

That may be news to users of computer algebra systems such as *Mathematica*. These systems, which manipulate symbolic expressions and do "exact" arithmetic, generally include a primality test based on one of the jazzed-up versions of Fermat's Little Theorem. Again, when one of these tests says a hundred- or thousand-digit number is composite, the result is reliable (even though, paradoxically, the test doesn't contain any clue as to what the factors might be!), but when the test says "prime," it really means "probably prime."

"It's not generally appreciated that these tests are not proofs of primality," says Pomerance. The chance of being burned by a prime impostor is pretty low—legitimate primes are far more common than Carmichael numbers—but it's still worth reading the results with a *caveat calculator*.

The Georgia trio's proof that Carmichael numbers pop up infinitely often is based on a heuristic argument put forward by Paul Erdős in 1956. The main idea is to choose a number L for which there are a large number of primes p that don't themselves divide L, but have the property that $p - 1$ divides L. The key point is then to show that these primes can be multiplied together in lots of different ways so that the products all leave remainder 1 when divided by L. It turns out that each such product is a Carmichael number.

For example, with $L = 120$, the primes in question are 7, 11, 13, 31, 41, and 61. A check of all possible combinations reveals that $41041 = 7 \times 11 \times 13 \times 41$, $172081 = 7 \times 13 \times 31 \times 61$, and $852841 = 11 \times 31 \times 41 \times 61$ all leave remainder 1 when divided by 120, and hence are Carmichael numbers.

The fact that the numbers constructed by Erdős's argument are Carmichael

"And then here I was, I had done all of this thinking about how to do it. . . and I was crushed, I was really crushed," says Alford. "So I went home that night, really with my nose just plain flat out of joint. But the next morning I woke up, and I knew how to construct 2^{100} of them."

numbers goes back to the mathematician A. Korselt. Korselt proved that a number n divides all numbers of the form $a^n - a$ (that is, n is a Carmichael number) if and only if it is squarefree (which means no prime divides it more than once) and has the property that $p - 1$ divides $n - 1$ whenever p is a prime divisor of n. Interestingly, he proved this in 1899—more than a decade before Carmichael put his stamp on the subject. The difference is, Korselt seemed to think there were no such numbers. They "surely would have been known as Korselt numbers had he just done a few computations!" says Granville.

What Alford, Granville, and Pomerance did, in essence, was to make Erdős's argument precise. Their work was spurred by recent work of Zhang Mingzhi at Sichuan University, who used the Erdős heuristic to produce examples of large Carmichael numbers. Alford, who is mainly a computational number theorist, thought he could do better. "I was champing at the bit to construct one a million digits long," he recalls. Instead, Pomerance challenged him to show that his method would produce huge families of such numbers.

"He said, 'Red, if it's as easy as you say it is, why don't you see if you can't construct 2^{50} of them,' " Alford recalls. "And then here I was, I had done all of this thinking about how to do it. . . and I was crushed, I was really crushed, for Carl to say it's just like writing down a bigger integer!" Alford laughs, then continues: "So I went home that night, really with my nose just plain flat out of joint. But the next morning I woke up, and I knew how to construct 2^{100} of them."

In fact, Alford's program coughed up 2^{128} Carmichael numbers (or, rather, prime numbers that can be combined in that many ways to produce Carmichael numbers). That enticed Granville and Pomerance to tackle the theoretic end of Erdős's argument. By using deep results in analytic number theory and combinatorial techniques from the theory of groups, they were finally able to flesh out Erdős's argument enough to prove a succinct theorem: There are more than $x^{2/7}$ Carmichael numbers up to x, for all sufficiently large x.

Exactly how large x has to be to be "sufficiently large" is still unclear (the analytic techniques are too convoluted to produce an estimate), although the numerical evidence suggests it happens at around $x = 10^7$. In a way it doesn't matter much, because Erdős's argument actually implies that the exponent $2/7$ can be replaced by any value short of 1. In other words, if you go far enough out, Carmichael numbers are amazingly abundant. The impostors are not as numerous as the true primes, but there are enough of them to make you stop and wonder.

An Imposing New Prime

In the early months of 1992—about the same time the Georgia trio proved the infinitude of prime impostors—a group at AEA Technology's Harwell Laboratory in Britain announced the discovery of a new "largest" prime number: a monster with nearly a quarter of a million digits belonging to a class of numbers known as "Mersenne primes."

The new number is not, of course, the largest possible prime. There is no such thing. What "largest" means here is it's the largest number *known* to be prime.

Mersenne primes all have the form $2^p - 1$, where the exponent p is itself a prime. If all primes p produced Mersenne primes, there'd be no sport in finding "largest" primes, but that's not the case. Mersenne primes are rare enough that only thirty-two of them are known to exist. The one found last year occurs for the exponent $p = 756,839$. A Cray-2 supercomputer took nineteen hours, using a variant of Fermat's Little Theorem known as Lucas's Test, to verify that the number indeed is prime.

Map-Coloring Theorists Look at New Worlds

People often think that once a "hard" mathematical problem has been solved, that's the end of the story. Nothing could be further from the truth. Mathematical problems rarely exist in a vacuum; the best ones are usually surrounded by a coterie of other interesting problems. Rather than spelling the end to a subject, the solution to a challenging problem more often means that researchers will turn with increased interest to some of the questions related to it.

Take the Four Color Theorem, for example. In 1976, Kenneth Appel and Wolfgang Haken at the University of Illinois proved that four colors are enough to paint any conceivable map in the plane in such a way that no countries with a common border are painted the same color. This breakthrough, which the researchers summarized as "four colors suffice," was one of the most talked-about results of the 1970s, in part because much of the proof was done on a computer. To many, the story of map coloring seemed over and done with.

Not so. Recently, researchers have been looking at the map-coloring problem for classes of maps to which the Four Color Theorem does not apply. These maps are drawn not on a flat piece of paper, but on arbitrary surfaces with any number of "handles" on them, such as a coffee cup or a two-handled vase. What the researchers have found can be stated as a nice counterpart to Appel and Haken's result: For these new classes of maps, *five* colors suffice.

Actually, there are two separate five-color theorems. Carsten Thomassen, a mathematician at the Technical University of Denmark, has proved that five colors are enough for maps on these many-handled surfaces, provided the countries to be colored are sufficiently small and numerous. Meanwhile, Neil Robertson at Ohio State University, Paul Seymour at Bell Communications Research (Bellcore) in Morristown, New Jersey, and Robin Thomas at Georgia Institute of Technology have reached the same conclusion for a different class of maps: Five colors suffice provided the countries to be colored can't be aligned into six mutually neighboring "federations."

The provisos of the two five-coloring theorems are poles apart, Robertson points out. That doesn't mean the two theorems are in conflict, though. Quite the contrary, it means the results combine to account for a large class of maps drawn on general surfaces.

Both new theorems belong to a branch of mathematics called graph theory. A mathematical graph is an extremely simple object: just a bunch of points (called *vertices*) with a bunch of curves (called *edges*) connecting them. Few things are simpler than that, yet few things lead so quickly to complicated problems and intricate results. Graph theory has come to play a key role in theoretical computer science and numerous other applied areas ranging from the design of transportation networks to the mathematics of chemical compounds (see box on page 46). Map coloring is just one area where graph theory plays a unifying role.

It's very easy to turn a map into a graph. You put a vertex at the capital of each country, and draw an edge between two vertices if the corresponding countries have a common border (see Figure 1). The graphs that result from ordinary maps

Figure 1. *Any map on the plane can be converted into a planar graph, and vice versa. Here, state capitals are connected by lines to form a graph.*

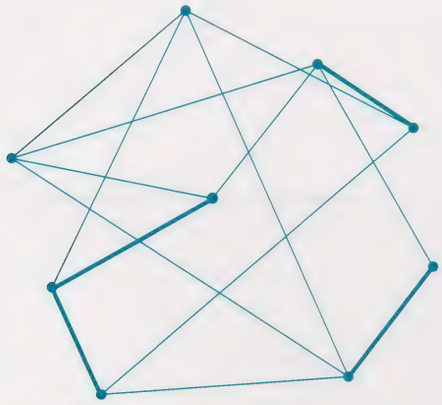


Figure 2. *If the dark edges are contracted so that their vertices merge, the resulting graph (called a "minor") is identical to K_5. Therefore this graph cannot be drawn on the plane without at least two edges crossing.*

are called "planar" graphs, because they can be drawn on a flat plane without any of their edges crossing. Because each vertex corresponds to a country, "coloring" a graph means coloring the vertices; if two vertices are connected by an edge, they should be given different colors. The idea that the edges do not cross corresponds to the fact that if two countries have a common border, you can make a road between the two capitals that travels only in those two countries.

The Four Color Theorem says that you need at most four different colors to color any planar graph. But not every graph is planar. In particular, the graph with five mutually adjacent vertices cannot be drawn in the plane without two edges crossing. (If it could, the Four Color Theorem would be false!) Another example is the "bipartite" graph consisting of two groups of three vertices, with an edge connecting each vertex in one group to each vertex in the other.

In 1930, the Polish mathematician Kazimir Kuratowski proved that these two graphs (usually denoted as K_5 and $K_{3,3}$) are essentially the only nonplanar graphs. What this means is that any other nonplanar graph contains at least one of these two, possibly in the form of a "minor," which is the graph theorist's term for a set of federations. (A federation forms when two neighboring countries erase their common border. For graphs, this amounts to contracting an edge until the vertices it connects merge. See Figure 2.) In other words, there are essentially only two "obstructions" to a graph being planar: K_5 and $K_{3,3}$.

While K_5 and $K_{3,3}$ cannot be drawn on a flat piece of paper, they can be drawn on a surface with a handle (see Figure 3). And indeed, every graph can be drawn on some surface with some number of handles, although determining exactly how many handles are necessary is not easy—in 1988 Thomassen proved that task to be NP-complete. (For an explanation of NP-complete problems, see "New Computer Insights from 'Transparent' Proofs," pages 7-11.)

In 1983, Robertson and Seymour generalized Kuratowski's theorem from the plane to surfaces with handles. They proved that for each such surface the set of "obstructing" graphs, while possibly quite large, is nevertheless always finite.

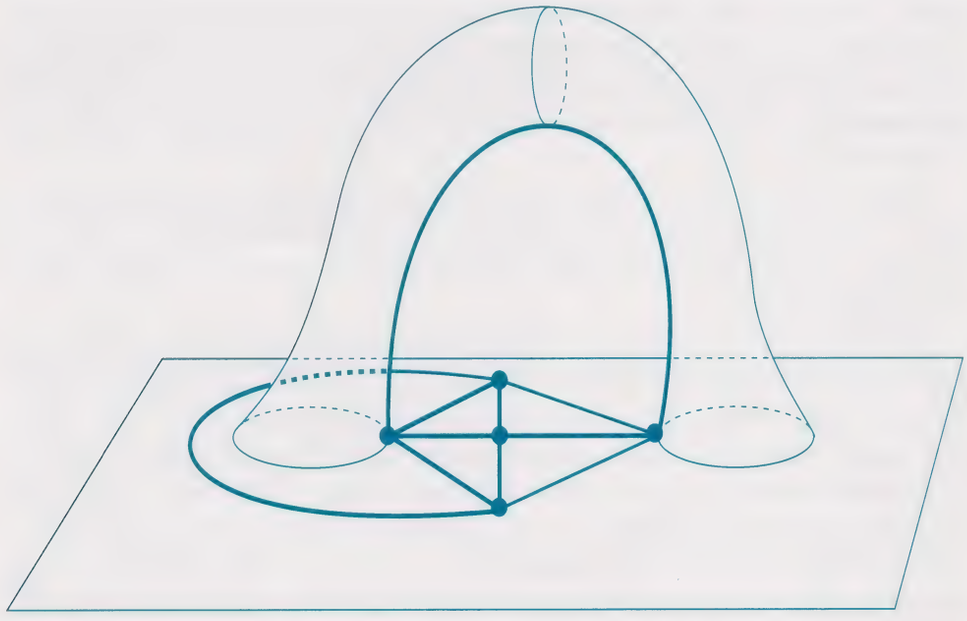


Figure 3. *K_5 can't be drawn on a plane without at least two edges crossing. But it can be drawn on a surface with a handle.*

Five Colors *Don't* Suffice!

In 1890, Percy Heawood proved that any map drawn on flat paper can be colored with no more than five different colors. Heawood's five-color theorem also applies to maps drawn on the globe, because, topologically speaking, the sphere and the plane are equivalent. However, it doesn't hold for graphs drawn on surfaces with handles. For example, it's possible to divide the torus into seven regions, each of which borders the other six, thus necessitating a separate color for each region.

How many colors does a surface with handles require? Heawood proved it never takes more than $(7 + \sqrt{48g + 1})/2$ colors, where g is the number of handles. In 1968, Gerhardt Ringel and J. W. T. Youngs proved that Heawood's bound (rounded down to the nearest integer) is exact: Every surface has maps that require as many colors as allowed by the formula.

Curiously, Heawood's formula gives the right answer—four—for the plane ($g = 0$), but the proof only works when the surface actually has handles. This is just one of many instances in mathematics where a problem is easier to solve in a complicated-sounding setting than it is in its original guise.

The Four Color Theorem says that you need at most four different colors to color every planar graph. But not every graph is planar.

This was an early result in an ongoing research program aimed at developing a "structural" theory of graphs.

Their latest work with Thomas is in the same vein. In 1943, Hugo Hadwiger conjectured that any graph can be colored with n colors provided it doesn't contain K_{n+1} (the analog of K_5, with $n + 1$ mutually adjacent vertices) as a minor. (Obviously if a graph contains K_{n+1} outright, there's no way to color it with just n colors. If K_{n+1} is present as a minor, it may still be possible to get by with n or even fewer colors, as in Figure 4, but that's not what Hadwiger's conjecture is concerned with. His conjecture is concerned with those graphs that don't contain K_{n+1} in any way, shape, or form.) In other words, Hadwiger's conjecture says the only potential obstruction to n-coloring a graph is the presence of K_{n+1}.

Hadwiger's conjecture is clearly true for $n = 1$ and 2, and easy to prove for $n = 3$. For $n = 4$, it turns out to be equivalent to the Four Color Theorem (although the equivalence is by no means easy to prove). Robertson, Seymour, and Thomas's theorem settles the case $n = 5$. Of course, that leaves an infinite amount of Hadwiger's conjecture unsolved. But "it says we're beginning to learn what's going on," Robertson notes.

Hadwiger's conjecture really doesn't care much what kind of surface a graph or its associated map is drawn on. Thomassen's five-color theorem, however, does. The problem it solves has a somewhat shorter history. In 1982, Michael Albertson at Smith College in Northampton, Massachusetts, and Walt Stromquist, a mathematician at Dan Wagner Associates in Paoli, Pennsylvania, proved that any map drawn on a torus (that is, a surface with one handle, like a coffee cup or its topological cousin, the doughnut) can be five-colored provided that any "tour" that travels all the way around the handle in any direction visits at least eight different countries.

Albertson and Stromquist conjectured that something similar should be true for surfaces with more handles: that if all trips around the handles are sufficiently long (that is, visit sufficiently many countries), then the map (or graph) should be five-colorable. That condition is known as "local planarity."

At one point it was thought that locally planar maps might even be four-colorable. However, Steve Fisk at Bowdoin College in Brunswick, Maine, put an end to that in 1978, by showing how to draw maps on the torus (or any other

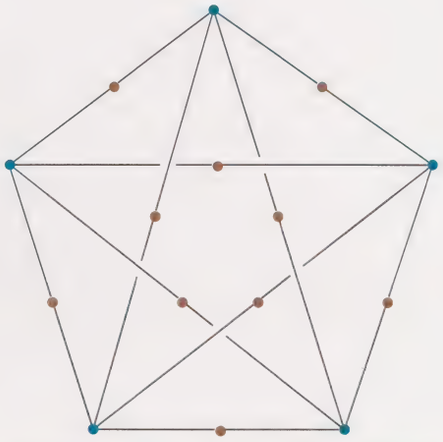


Figure 4. *This graph has been two-colored even though it contains K_5 as a minor. Does that violate Hadwiger's conjecture? No!*

Fan Chung and Shlomo Sternberg with a model of a buckyball.

surface with handles) with countries as small as you please that nevertheless require five colors.

In 1985, Joan Hutchinson, now at Macalester College in St. Paul, Minnesota, proved that five colors suffice if "small" is defined in a particular way. Thomassen's theorem goes further. It says that for each surface, there is a number n such that if all trips that go around handles visit at least n countries of a given map, then that map is five-colorable. The number n depends on the number of handles on the surface; Thomassen's proof provides only an estimate, which doubles each time another handle is added. The actual number, he notes, is likely to be a good deal less.

One may well ask, why bother? One answer is that graph coloring is not a purely academic exercise; it does have applications in real settings. "What graph coloring is really about is scheduling," says Stromquist. "Trying to color a graph is the same as trying to schedule a whole lot of events into time slots in such a way that incompatible events don't happen at the same time." Researchers interested in developing algorithms for scheduling consequently often find themselves faced with problems in graph coloring.

But the main reason continues to be one of pure intellectual challenge. The proof of the Four Color Theorem still requires a computer to rule out a thousand or so ways a planar map might require five colors rather than four, and to many mathematicians that's an unsatisfactory state of affairs. "There's obviously still something to be learned about graph coloring," says Robertson. Theorists believe that the results of research in map coloring will be of use in other areas of graph theory and its applications. Beyond that, there's one final, inarguable reason, summarized succinctly by Stromquist: "It's fun."

Graph Theory Tackles the Buckyball

Graph theory is one of the most playful topics in mathematics. But it's also one of the most useful. In part because of the way it combines algebraic abstractions with down-to-earth geometric configurations, graph theory turns up in places you might not expect—and graph theorists often wind up working in areas seemingly far afield.

Take Fan Chung, for example. Chung, a mathematician at Bellcore, is an expert on graph invariants. Normally she works either in pure theory or on applications to problems in communication networks. But recently she has been applying her graph-theoretic expertise to something quite different: buckyballs.

First discovered just a few years ago, the buckyball is a soccerball-shaped molecule consisting of sixty carbon atoms arranged in a highly symmetric, icosahedral pattern. An outline of the buckyball's chemical bonds is reminiscent of the "geodesic domes" popularized in the 1960s by R. Buckminster Fuller—hence the official chemical name, *buckminsterfullerene*.

Chemists, physicists, and materials scientists have flocked to the buckyball and its variants like moths drawn to candlelight (indeed, the soot that rises with the flame of a candle may consist in part of buckyballs). The crowd now includes a handful of mathematicians.

Graph theory is no newcomer to chemistry: The term "graph" was first used (in its technical sense) by the mathematician J. J. Sylvester in 1877, in a paper titled "Chemistry and Algebra." Mathematicians' drawings of graphs and chemists' renderings of chemical compounds are strikingly similar. That's no coincidence. Graph theory turns out to be a useful mathematical tool for chemists seeking to understand the myriad ways that elements can combine to form complex molecules and the myriad properties those molecules may possess. *(continued on next page)*

Chung has been working with Shlomo Sternberg at Harvard University to analyze the mathematical properties of the buckyball's unique structure. Sternberg is an expert in the theory of group representations, which can be roughly described as the systematic study of symmetry. The combination of graph theory and representation theory, applied to the buckyball, is a powerful one. Chung and Sternberg's analysis so far has gone a long way toward explaining the buckyball's spectroscopic properties and why the molecule is so stable.

Chung and Sternberg are also trying to find a theoretical explanation of another property of the buckyball, which it shares with a growing family of promising new materials, namely high-temperature superconductivity. The discovery of high-temperature superconductivity in the late 1980s has left theorists scrambling for explanations of a phenomenon that isn't even well understood for low temperatures. "It's really a very big puzzle," says Chung. However, the fact that so many materials turn out to be superconducting suggests that the puzzle may not depend on details of the physics. "It's our belief it must be a mathematical explanation," Chung concludes.

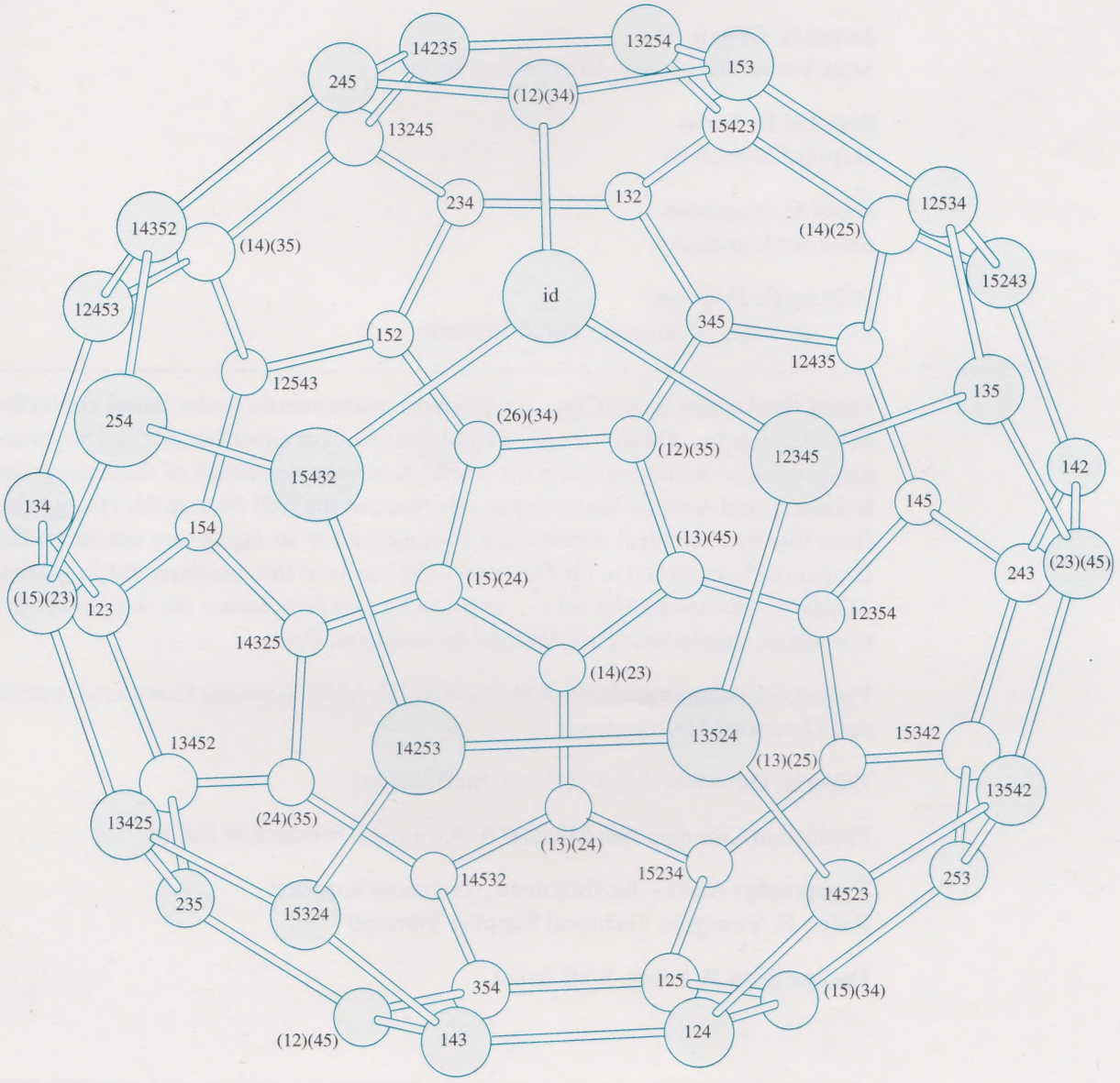


Figure 5. *The Chung-Sternberg buckyball. (Based on figure courtesy of Fan R. K. Chung.)*

Credits

About the Author: Barry Cipra is a freelance mathematics writer based in Northfield, Minnesota. He is currently a Contributing Correspondent for *Science* magazine and also writes regularly for *SIAM News*, the newsletter of the Society for Industrial and Applied Mathematics. He received the 1991 Merten M. Hasse Prize from the Mathematical Association of America for an expository article on the Ising model, published in the December 1987 issue of the *American Mathematical Monthly*. His book, *Misteaks ... and how to find them before the teacher does...* (a calculus supplement), is published by Academic Press.

Project Administration: Samuel M. Rankin, III, AMS Associate Executive Director and Director of Publications

Editorial Direction: Allyn Jackson, Staff Writer

Production Coordination: Michael A. Kowalski, Production Editor

Typography: Neil G. Bartholomew, Technical Support;
Ralph E. Youngen, Technical Support Manager

Design: Peter B. Sykes, Staff Artist

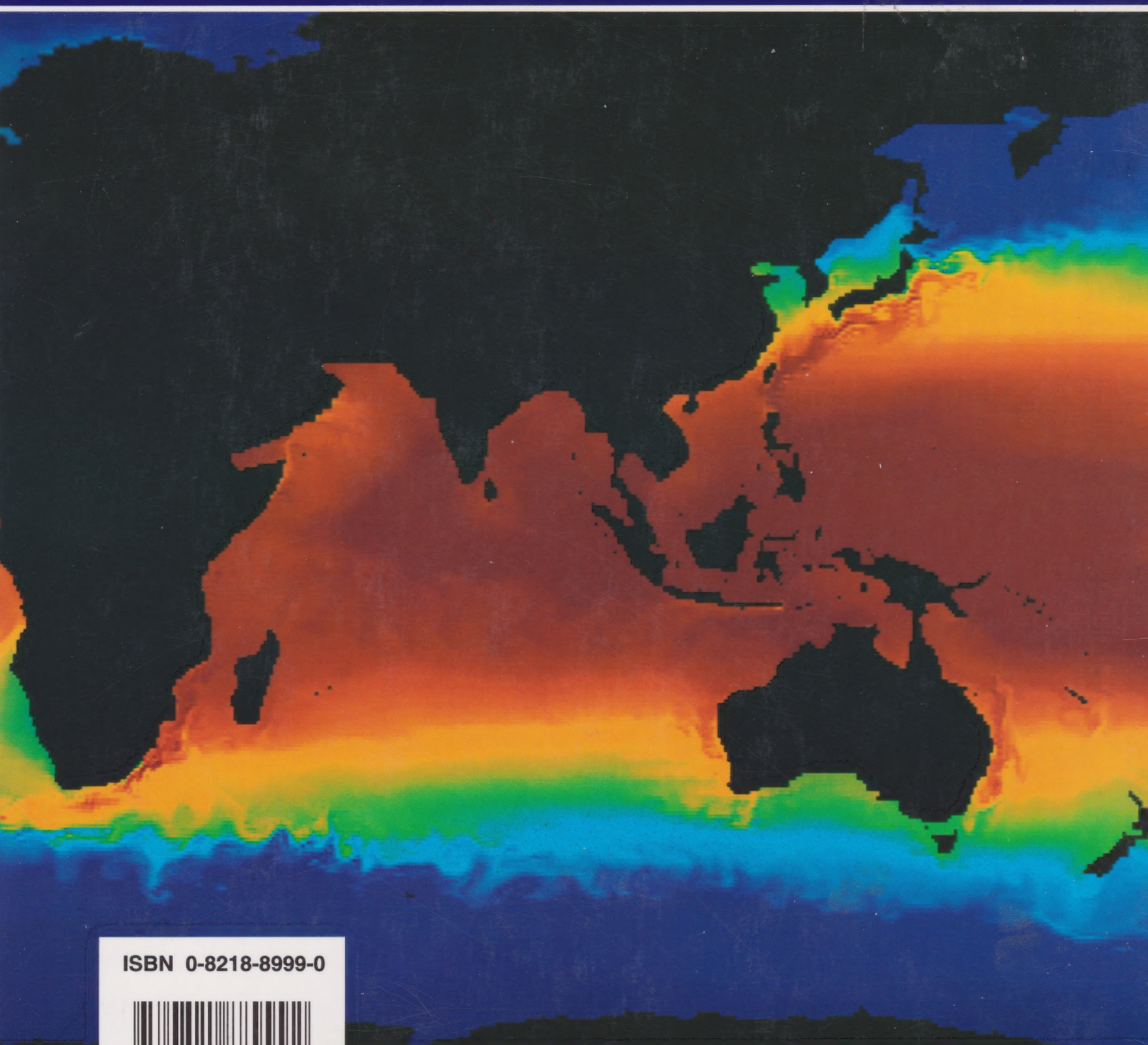
The AMS gratefully acknowledges the support of the Exxon Education Foundation and the Alfred P. Sloan Foundation for the publication and distribution of *What's Happening in the Mathematical Sciences*.

About the American Mathematical Society

The American Mathematical Society is a nonprofit organization devoted to research in the mathematical sciences. For more than 100 years, the Society has worked to support the advance of mathematical research, the communication of mathematical ideas, and the improvement of the mathematics profession. In recent years, the Society has increased its attention to mathematics education, public awareness of mathematics, and the connections of mathematics research to its uses.

The AMS is the world's largest mathematical organization, with nearly 30,000 members. As one of the world's major publishers of mathematical literature, the Society produces a wide range of book series, journals, monographs, and videotapes, as well as the authoritative reference, *Mathematical Reviews*. The Society is a world leader in the use of computer technology in publishing and is involved in the development of electronic means of information delivery. Another primary activity of the AMS is organizing meetings and conferences. In addition to the annual winter meeting, the Society organizes bi-annual summer meetings, as well as numerous smaller meetings during the academic year and workshops, symposia, and institutes during the summer. Other major Society activities are employment services, collection of data about the mathematical community, and advocacy for the discipline and the profession.

The main headquarters of the Society, located in Providence, Rhode Island, employs nearly 200 people and contains a large computer system, a full publication facility, and a warehouse. Approximately eighty people are employed at the *Mathematical Reviews* office in Ann Arbor, Michigan. In September 1992 the AMS opened an office in Washington, DC, in order to enhance the Society's public awareness efforts and its linkages with federal science policy.



ISBN 0-8218-8999-0

9 780821 889992